

Inversion start site

```
ATTATAAAGGAAAAAGAAAATAACGCAATGGACAAGTGGTG
860 +-----+-----+-----+-----+ 900
TAATATTTCTTTTCTTTTATTGCGTTACCTGTTTACCAC (41)
    Y K G K R K * R N G Q V V

AAGCTGTGAAGTCTCAGGTGTGCACAATTATCAGGAACA TCCCAAACCAAAGTGAGGTAGA
901 -----+-----+-----+-----+-----+ 960
TTCGACACTTGAGTCCACACGTGTTAATAGTCCTTGTGGGGTTTGGTTTCACTCCATCT (101)
    K L * T Q V C T I I R N T P K P K * G R

AATAGCATGAGAAGCCGTGTTTGATGTTAATTAATT
961 -----+-----+----- 996
TTATCGTACTGTTTCGGGCACAACTACAATTAATTAA (137)
    N S M R S R V * C * L I
```

The inversion sequence of the apo-dystrophin-4 cDNA (SEQ ID NO 1)

Figure 1

10956264.092801

0996364-09304

Inversion start site  
|  
TAAAGAAAGAATTATAAAGGAAAAAGAAAATAACGCAATGGACAAGTGGTG  
850 +-----+-----+-----+-----+-----+ 900  
ATTTCTTTCTTAATATTTCTTTTCTTTTATTGCGTTACCTGTTCAACAC (51)  
\* R K N Y K G K R K \* R N G Q V V  
  
AAGCTGTGAAGCTCAGGTGTGCACAATTATCAGGAACACCCCAAAGTGAGGTAGA  
901 -----+-----+-----+-----+-----+ 960  
TTCGACACTTGAGTCCACACGTGTTAATAGTCCTTGTGGGGTTTGGTTTCACTCCATCT (111)  
K L \* T Q V C T I I R N T P K P K \* G R  
  
AATAGCATGAGAAGCCGTGTTTGATGTTAATTAATT  
961 -----+-----+-----+-----+----- 996  
TTATCGTACTCTTCGGCACAACTACAATTAATTAA (147)  
N S M R S R V \* C \* L I

The inversion sequence of the apo-dystrophin-4 cDNA plus a 10 base-pair region 5' to the start of the inversion sequence (SEQ ID NO 1A).

Figure 1A

0966264.092601

```

Start at 710
|
AACAAATGGCAG
+-----+ 720
TTGTTACCGTC (11)
  Q W Q

721 GTTTTACACGTCTATGCAATTGTACAAAAAGTTATAAGAAAACCTACATGTAAATCTTG
-----+-----+-----+-----+-----+-----+ 780
CAAATGTGCAGATACGTTAACATGTTTTTTCAATATCTTTTGATGTACATTTTAGAAC (71)
V L H V Y A I V Q K S Y K K T T C K I L

781 ATAGCTAAATAACTTGCCATTTCTTTATATGGAACGCATTTTGGGTGTTTAAAAATTTA
-----+-----+-----+-----+-----+-----+ 840
TATCGATTTATTTGAACGGTAAAGAAATATACCTTGCGTAAAACCCAACAATTTTAAAT (131)
I A K * L A I S L Y G T H F G L F K N L
      inversion start site
      |
841 TAACAGTTATAAAGAAAGAATTATAAAGGAAAAAGAAAATAACGCAATGGACAAGTGGTG
-----+-----+-----+-----+-----+-----+ 900
ATTGTCAATATTTCTTTCTTAATATTTCTTTTCTTTTATTGCGTTACCTGTTTACCAC (191)
* Q L * R K N Y K G K R K * R N G Q V V

901 AAGCTGTGAACTCAGGTGTGCACAATTATCAGGAACACCCCAAAACCAAAGTGAGGTAGA
-----+-----+-----+-----+-----+-----+ 960
TTCGACACTTGAGTCCACACGTGTTAATAGTCCTTGTGGGGTTTTGGTTTCACTCCATCT (251)
K L * T Q V C T I I R N T P K P K * G R

961 AATAGCATGAGAAGCCGTGTTTGATGTTAATTAATT
-----+-----+-----+-----+-----+ 996
TTATCGTACTCTTCGGCACAACTACAATTAATTAA (287)
N S M R S R V * C * L I

```

**The inversion sequence of the apo-dystrophin-4 cDNA plus the upstream 150 bp from the start of the inversion at 860 to the Hpa I enzyme site (SEQ ID NO 1B)**

**Figure 1B**

0996264.092301

```
GTGGTTTGATTGATAGTAAAAAATGTTTCGTTAATACAAGTAGAGAGTAAGTAATCAAT
1  -----+-----+-----+-----+-----+-----+ 60
CACCAAACAACTATCATTTTTTTTACAAGCAATTATGTTTCATCTCTCATTATTAGTTA
V V * L I V K K M F V N T S R E * V I N

CAATCACTCATAGCCAAGGTGGAAAAGATGTATCCCATCATGGAATATTCCTGTTCTGAT
61 -----+-----+-----+-----+-----+-----+ 120
GTTAGTGAGTATCGGTTCCACCTTTTCTACATAGGGTAGTACCTTATAAGGACAAGACTA
Q S L I A K V E K M Y P I M E Y S C S D

AGAAATCTTGTGCTTATCTATGGAATTCTTTTGATATATATTTACATTGGGAACCTGAAT
121 -----+-----+-----+-----+-----+-----+ 180
TCTTTAGAACACGAATAGATACCTTAAGAAAACATATATATAAATGTAACCTTGGACTTA
R N L V L I Y G I L L I Y I Y I G N L N

GTAGCTTGACATTTTCCATGTAAACACCAGTAGCCTGATCCAACATTAAGCTGATACTA
181 -----+-----+-----+-----+-----+-----+ 240
CATCGAACTGTAAAAAGGTACATTTGTGGTCATCGGACTAGGTTGTAATTCGACTATGAT
V A * H F S M * T P V A * S N I K L I L

ACAAACAACGTGTAATGGCTTCATTAATAAGGCTTTGCTTCTTCCTGGAAACTGGTGAAA
241 -----+-----+-----+-----+-----+-----+ 300
TGTTTGTGTCACATTACCGAAGTAATTATCCGAAACGAAGAAGGACCTTTGACCACTTT
T N N V * W L H * * G F A S S W K L V K

AATCAAACCTTGTTGTGTACACCCTCGATGCAGCTTCTGTGTTGTCTTCACCCAGAAATG
301 -----+-----+-----+-----+-----+-----+ 360
TTAGTTTGGAACAACACATGTGGGAGCTACGTCGAAGACACAACAGAAGTGGGTCTTTAC
N Q T L L C T P S M Q L L C C L H P E M
```

The polynucleotide sequence of apo-dystrophin-4 (SEQ ID NO 2)

Figure 2

```

GGGAATGATTTCCCAAATGGCAAAGAAACAGAGTGATGCTATCTATCTGCACCTTTTGTG
361 -----+-----+-----+-----+-----+-----+ 420
CCCTTACTAAAGGGTTTACCGTTTCTTTGTCTCACTACGATAGATAGACGTGGAAAACAT
G N D F P N G K E T E * C Y L S A P F V

AAGTCTGTCTTTCTTTCTCTTTGTTTTCCAGGACACAATGTAGGAAGTCTTTTCCACATG
421 -----+-----+-----+-----+-----+-----+ 480
TTCAGACAGAAAAGAAAGAGAAACAAAAGGTCCTGTGTTACATCCTTCAGAAAAGGTGTAC
K S V F L S L C F P G H N V G S L F H M

GCAGATGATTTGGGCAGAGCGATGGAGTCCTTAGTATCAGTCATGACAGATGAAGAAGGA
481 -----+-----+-----+-----+-----+-----+ 540
CGTCTACTAAACCCGTCTCGCTACCTCAGGAATCATAGTCAGTACTGTCTACTTCTTCCT
A D D L G R A M E S L V S V M T D E E G

GCAGAATAAATGTTTTACAACCTCCTGATTCCCGCATGGTTTTTATAATATTCATACAACA
541 -----+-----+-----+-----+-----+-----+ 600
CGTCTTATTTACAAAATGTTGAGGACTAAGGGCGTACCAAAAATATTATAAGTATGTTGT
A E * M F Y N S * F P H G F Y N I H T T

AAGAGGATTAGACAGTAAGAGTTTACAAGAAATAAATCTATATTTTGTGAAGGGTAGTG
601 -----+-----+-----+-----+-----+-----+ 660
TTCTCCTAATCTGTCATTCTCAAATGTTCTTTATTTAGATATAAAAACACTTCCCATCAC
K R I R Q * E F T R N K S I F L * R V V

GTATTATACTGTAGATTTTCAGTAGTTTCTAAGTCTGTTAT'TGTTTGTTAACAATGGCAG
661 -----+-----+-----+-----+-----+-----+ 720
CATAATATGACATCTAAAGTCATCAAAGATTCAGACAATAACAAAACAATTGTTACCGTC
V L Y C R F Q * F L S L L L F C * Q W Q

```

Figure 2 (cont'd)

09966264.092801

GTTTTACACGTCTATGCAATTGTACAAAAAAGTTATAAGAAAACCTACATGTAAAATCTTG  
721 -----+-----+-----+-----+-----+-----+ 780  
CAAAATGTGCAGATACGTTAACATGTTTTTTCAATATTCTTTTGATGTACATTTTAGAAC  
V L H V Y A I V Q K S Y K K T T C K I L  
ATAGCTAAATAACTTGCCATTTCTTTATATGGAACGCATTTTGGGTTGTTTAAAAATTTA  
781 -----+-----+-----+-----+-----+-----+ 840  
TATCGATTTATTGAACGGTAAAGAAATATACCTTGCGTAAAACCCAACAAATTTTAAAT  
I A K \* L A I S L Y G T H F G L F K N L  
TAACAGTTATAAAGAAAGAATTATAAAGGAAAAAGAAAATAACGCAATGGACAAGTGGTG  
841 -----+-----+-----+-----+-----+-----+ 900  
ATTGTCAATATTTCTTTCTTAATATTCCTTTTCTTTTATTGCGTTACCTGTTCAACCAC  
\* Q L \* R K N Y K G K R K \* R N G Q V V  
AAGCTGTGAACTCAGGTGTGCACAATTATCAGGAACACCCCAAACCAAGTGAGGTAGA  
901 -----+-----+-----+-----+-----+-----+ 960  
TTCGACACTTGAGTCCACACGTGTTAATAGTCCTTGTGGGGTTTTGGTTTCACTCCATCT  
K L \* T Q V C T I I R N T P K P K \* G R  
AATAGCATGAGAAGCCGTGTTTGATGTTAATTAATT  
961 -----+-----+-----+----- 996  
TTATCGTACTCTTCGGCACAACTACAATTAATTAA  
N S M R S R V \* C \* L I

Figure 2 (cont'd)

09966264-092801

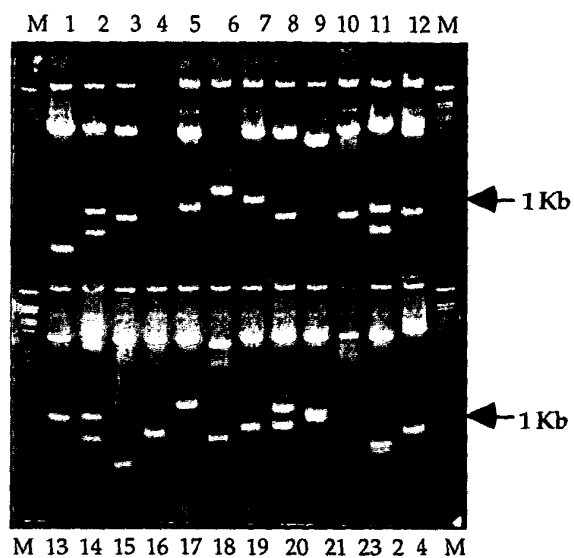


Figure 3A

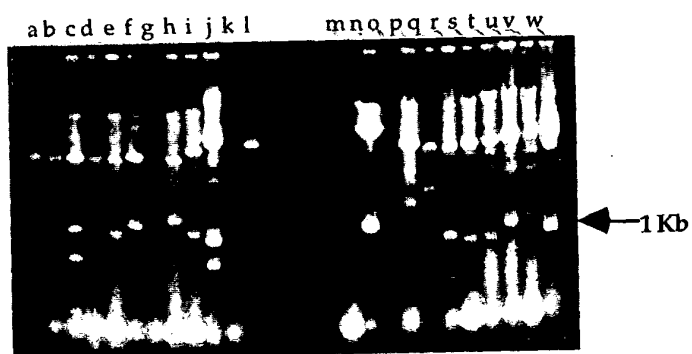


Figure 3B

099664-099966D

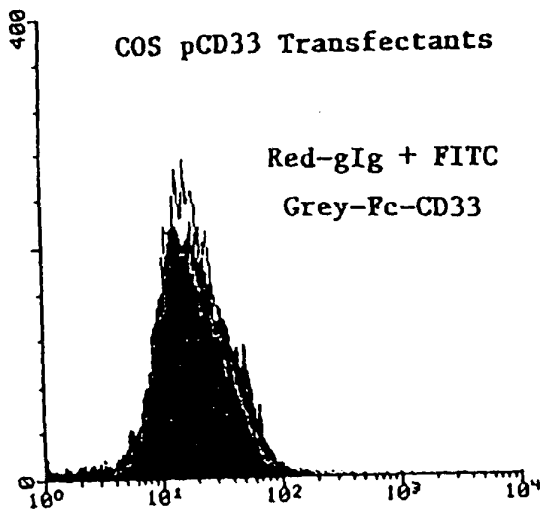


Figure 4A

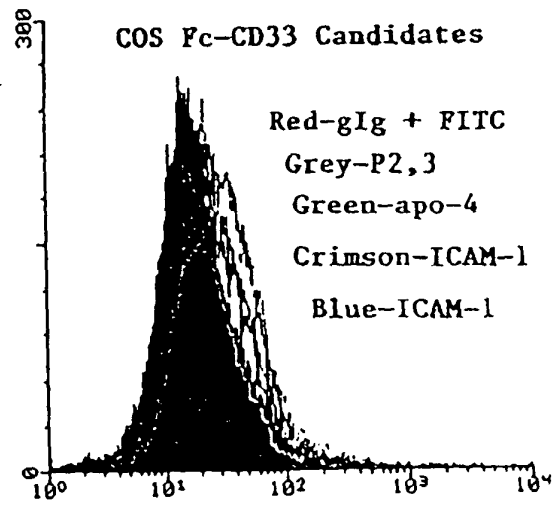


Figure 4B

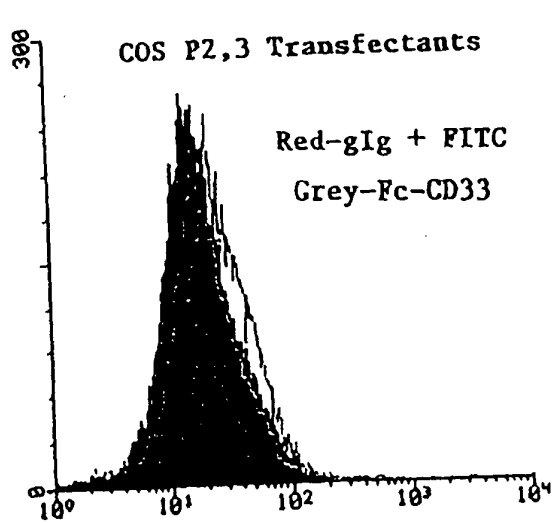
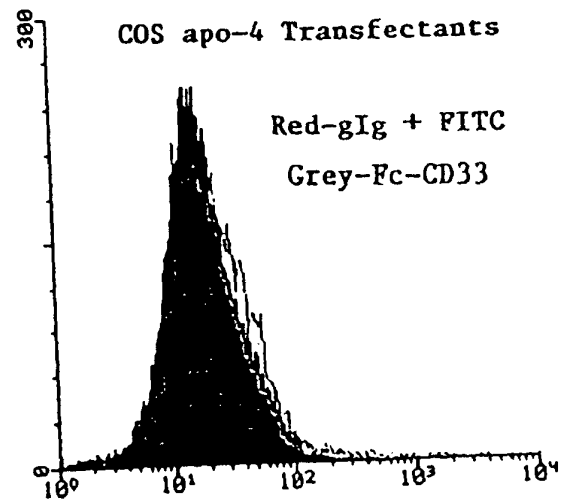


Figure 4C



Figur 4D



0996664-032801

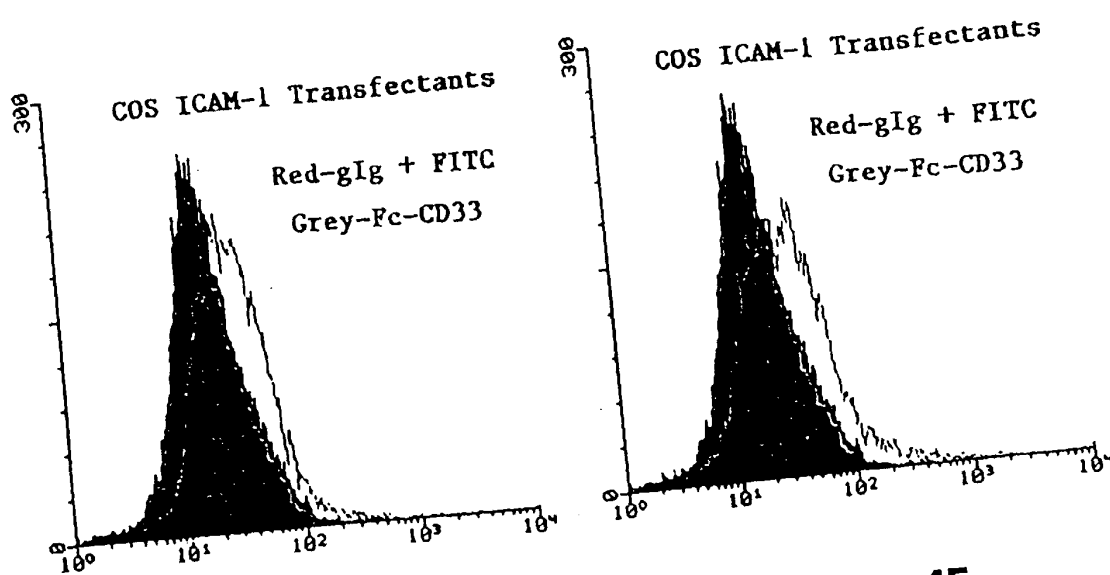


Figure 4E

Figure 4F

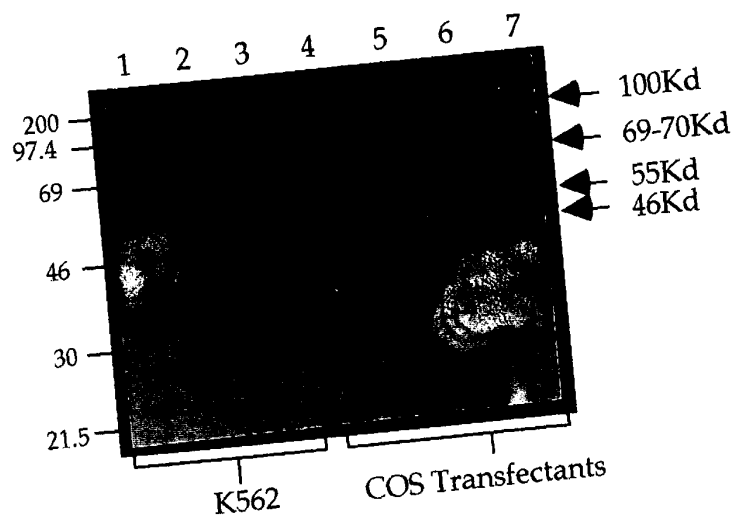


Figure 5

```

...TAGTTTCCTATTCAATGTATAGTGCACCAAAGGTCAATTCAAGAGTTTATTATTATT
-239 -----+-----+-----+-----+-----+-----+-----+
...ATCAAAGGATAAGTTACATATCACGTGGTTTCCAGTTAAGTTCTCAAATAATAATAA
. * F P I Q C I V H Q R S I Q E F I I I
-180

ATTTTCAACCCAAGTAAAAGCAGAGAGAAAATAGCCACCTCCACCATAGCCTCAGAAGCA
-179 -----+-----+-----+-----+-----+-----+-----+
TAAAAGTTGGGTTCATTTTCGTCTCTCTTTTATCGGTGGAGGTGGTATCGGAGTCTTCGT
I F N P S K S R E K I A T S T I A S E A
-120

AGCCAACAGCCTGAAACAGCTTTGAAATGAAAAGTTGGTGTGGCGGTGATGGTGGCAGTG
-119 -----+-----+-----+-----+-----+-----+-----+
TCGGTTGTGCGGACTTTGTGCGAACTTTACTTTTCAACCACACCGCCACTACCACCGTCAC
S Q Q P E T A L K * K V G V A V M V A V
-60

ATAATGGTGACCGATGGTTGGGTGCTGGTGTAGTGGTAGTTGTGAAGGTGGTGTATG
-59 -----+-----+-----+-----+-----+-----+-----+
TATTACCACTGGCTACCAACCCACGACCACTACCATCACCATCAACACTTCCACCACTAC
I M V T D G W V L V M V V V V V V K V V M
0

GTGTTTGTATTGATAGTAAAAAAATGTTCTGTTAATACAAGTAGAGAGTAAGTAATCAAT
1 -----+-----+-----+-----+-----+-----+-----+
CACCAAACTAACTATCATTTTTTTTACAAGCAATTATGTTTCATCTCTCATTTCATTAGTTA
V V * L I V K K M F V N T S R E * V I N
60

CAATCACTCATAGCCAAGGTGAAAAAGATGTATCCCATCATGGAATATTCCTGTTCTGAT
61 -----+-----+-----+-----+-----+-----+-----+
GTTAGTGAGTATCGGTTCCACCTTTTCTACATAGGGTAGTACCTTATAAGGACAAGACTA
Q S L I A K V E K M Y P I M E Y S C S D
120

AGAAATCTTGCTTATCTATGGAATTCTTTTGATATATATTTACATTGGGAACCTGAAT
121 -----+-----+-----+-----+-----+-----+-----+
TCTTTAGAACACGAATAGATACCTTAAGAAACTATATATAAATGTAACCCTTGGACTTA
R N L V L I Y G I L L I Y I Y I G N L N
180

GTAGCTTGACATTTTTCCATGTAAACACCAGTAGCCTGATCCAACATTAAGCTGATACTA
181 -----+-----+-----+-----+-----+-----+-----+
CATCGAACTGTAAAAAGGTACATTTGTGGTCATCGGACTAGGTTGTAATTCGACTATGAT
V A * H F S M * T P V A * S N I K L I L
240

ACAAACAACGTGTAATGGCTTCATTAATAAGGCTTTGCTTCTTCCTGGAAACTGGTGAAA
241 -----+-----+-----+-----+-----+-----+-----+
TGTTTGTGTGCACATTACCGAAGTAATTATCCGAAACGAAGAAGGACCTTTGACCACTTT
T N N V * W L H * * G F A S S W K L V K
300

AATCAAACCTTGTTGTGTACACCCTCGATGCAGCTTCTGTGTTGTCTTCACCCAGAAATG
301 -----+-----+-----+-----+-----+-----+-----+
TTAGTTTGGAAACAACACATGTGGGAGCTACGTGGAAGACACAACAGAAGTGGGTCTTTAC
N Q T L L C T P S M Q L L C C L H P E M
360

GGGAATGATTTCCCAAATGGCAAAGAAACAGAGTGATGCTATCTATCTGCACCTTTTGTA
361 -----+-----+-----+-----+-----+-----+-----+
CCCTTACTAAAGGGTTTACCGTTTCTTTGTCTCACTACGATAGATAGACGTGGAAAACAT
G N D F P N G K E T E * C Y L S A P F V
420

```

Figure 6

begin exon 79

|

421 AAGTCTGTCTTTCTTTCTTTGTTTTCCAGGACACAATGTAGGAAGTCTTTTCCACATG 480  
 -----+-----+-----+-----+-----+-----+-----+  
 TTCAGACAGAAAGAAAGAGAAACAAAAGGTCCTGTGTTACATCCTTCAGAAAAGGTGTAC  
 K S V F L S L C F P G H N V G S L F H M

481 GCAGATGATTTGGGCAGAGCGATGGAGTCCTTAGTATCAGTCATGACAGATGAAGAAGGA 540  
 -----+-----+-----+-----+-----+-----+-----+  
 CGTCTACTAAACCCGTCTCGCTACCTCAGGAATCATAGTCAGTACTGTCTACTTCTTCCT  
 A D D L G R A M E S L V S V M T D E E G

541 GCAGAATAAATGTTTTACAACCTCCTGATTCCCGCATGGTTTTTATAATATTCATACAACA 600  
 -----+-----+-----+-----+-----+-----+-----+  
 CGTCTTATTTACAAAATGTTGAGGACTAAGGGCGTACCAAAAATATTATAAGTATGTTGT  
 A E \* M F Y N S \* F P H G F Y N I H T T

(----N----)

601 AAGAGGATTAGACAGTAAGAGTTTACAAGAAATAAATCTATATTTTTGTGAAGGGTAGTG 660  
 -----+-----+-----+-----+-----+-----+-----+  
 TTCTCCTAATCTGTCATTCTCAAATGTTCTTTATTTAGATATAAAAACACTTCCCATCAC  
 K R I R Q \* E F T R N K S I F L \* R V V

661 GTATTATACTGTAGATTTCACTAGTTTCTAAGTCTGTTATTGTTTTGTTAACAATGGCAG 720  
 -----+-----+-----+-----+-----+-----+-----+  
 CATAATATGACATCTAAAGTCATCAAAGATTGAGACAATAACAAAACAATTGTTACCGTC  
 V L Y C R F Q \* F L S L L L F C \* Q W Q

721 GTTTTACACGTCTATGCAATTGTACAAAAAGTTATAAGAAAACCTACATGTAAAATCTTG 780  
 -----+-----+-----+-----+-----+-----+-----+  
 CAAAATGTGCAGATACGTTAACATGTTTTTTCAATATTCTTTTGATGTACATTTTAGAAC  
 V L H V Y A I V Q K S Y K K T T C K I L

781 ATAGCTAAATAACTTGCCATTTCTTTATATGGAACGCATTTTGGGTGTTTAAAAATTTA 840  
 -----+-----+-----+-----+-----+-----+-----+  
 TATCGATTTATTGAACGGTAAAGAAATATACCTTGCCTAAAACCCAACAAATTTTTAAAT  
 I A K \* L A I S L Y G T H F G L F K N L  
 inversion start site

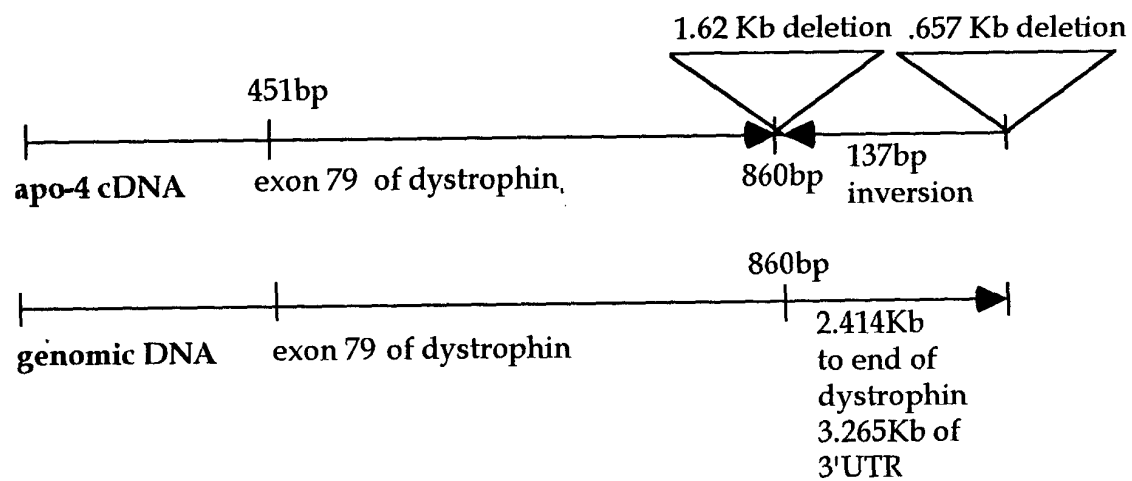
|

841 TAACAGTTATAAAGAAAGAATTATAAAGGAAAAAGAAAATAACGCAATGGACAAGTGGTG 900  
 -----+-----+-----+-----+-----+-----+-----+  
 ATTGTCAATATTTCTTTCTTAATATTTCTTTTCTTTTATTGCGTTACCTGTTCAACCAC  
 \* Q L \* R K N Y K G K R K \* R N G Q V V

901 AAGCTGTGAACTCAGGTGTGCACAATTATCAGGAACACCCCAAAACCAAGTGAGGTAGA 960  
 -----+-----+-----+-----+-----+-----+-----+  
 TTCGACACTTGAGTCCACACGTGTTAATAGTCCTTGTGGGGTTTGGTTTCACTCCATCT  
 K L \* T Q V C T I I R N T P K P K \* G R

961 AATAGCATGAGAAGCCGTGTTTGATGTTAATTAATT 996  
 -----+-----+-----+-----+-----+-----+-----+  
 TTATCGTACTCTTCGGCACAACTACAATTAATTAA  
 N S M R S R V \* C \* L I

Figure 6 (cont'd)

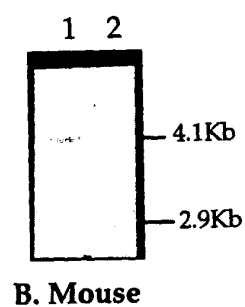
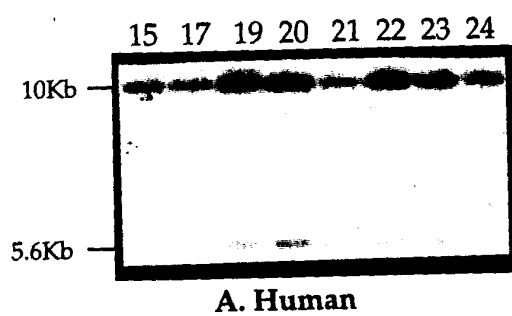
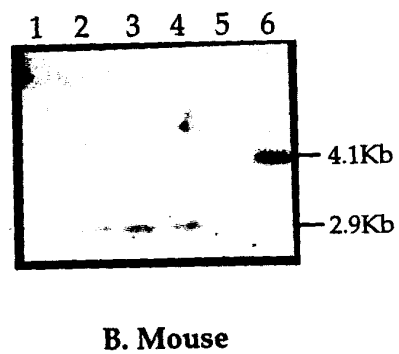
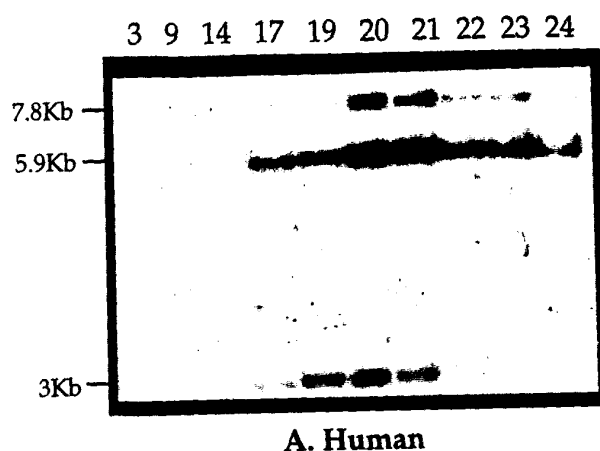


**Figure 7**

09966264-092301



**A. Human**



	50				1
Mgen1073	.....	.....	.....	.....	.....
Hapo1234	ctagtttcct	attcaatgta	tagtgcacca	aaggtcaatt	caagagttta
Consensus	-----	-----	-----	-----	-----
	51				100
Mgen1073	.....	.....	.....	.....	.....
Hapo1234	ttattattat	tttcaaccga	agtaaaagca	gagagaaaat	agccacctcc
Consensus	-----	-----	-----	-----	-----
	101		begin GRAIL exon @149 150		
Mgen1073	.....	.....	ttcACAGgCT	tAAgCAGCca	gtAAATGAcA
Hapo1234	accatagcct	cagaagcaag	ccaACAGcCT	gAAaCAGCtt	tgAAATGAaA
Consensus	-----	-----	---ACAG-CT	-AA-CAGC--	--AAATGA-A
	151				200
Mgen1073	AtT.....	.....	.....T	AtgtGgtAgt	cAgGtcactG
Hapo1234	AgTtggtgtg	gcggtgatgg	tggcagtgaT	AatgGtgAcc	gAtGgttggG
Consensus	A-T-----	-----	-----T	A---G--A--	-A-G-----G
	201			apo-4 5'end	250
Mgen1073	TGCTGGTaAT	GGTgaTctTA	GcaGgcAgAG	aaGGTGgTaG	TGaTTTGATa
Hapo1234	TGCTGGTgAT	GGTagTggTA	GttGtgA.AG	gtGGTGaTgG	TGgTTTGATt
Consensus	TGCTGGT-AT	GGT--T--TA	G--G--A-AG	--GGTG-T-G	TG-TTTGAT-
	251	M1			300
Mgen1073	GtaAaagtgt	AgAcTaTaCa	acAgaAtAAa	TAcAagtata	GTAA.....
Hapo1234	GatAgtaaaaa	AaAaTgTtCg	ttAatAcAAg	TAgAgagtaA	GTAAatcaatc
Consensus	G--A-----	A-A-T-T-C-	--A--A-AA-	TA-A-----A	GTAA-----
	301		M2	M3	350
Mgen1073	.....atc	caaCAAaGTG	tgAAAGgTGT	gTgCCATtAc	acAtctTTCT
Hapo1234	aatcactcat	agcCAAGGTG	gaAAAGaTGT	aTcCCATcAt	ggAataTTCc
Consensus	-----	---CAA-GTG	--AAAG-TGT	-T-CCAT-A-	--A---TTC-
	351				400
Mgen1073	cG.....	GtgATAagag	cTTgTCTAT	GaAgTTC...	TGAgATgTgT
Hapo1234	tGttctgata	GaaATcttgt	gCTTaTCTAT	GgAaTTTctt	TGAtATaTaT
Consensus	-G-----	G--AT-----	-CTT-TCTAT	G-A-TTC---	TGA-AT-T-T
	401				450
Mgen1073	TaggAagatG	AATcatcAat	TtaCaT....	TTcTcCCcat	cAAAtgaCac
Hapo1234	TtacAttggG	AACctgaAtg	TagCtTgaca	TTtTtCCatg	tAAAcacCAg
Consensus	T---A----G	AA-C---A--	T--C-T----	TT-T-CC---	-AAA---CA-
	451		begin mouse GRAIL exon		500
Mgen1073	cAtgCTGATC	CAGtATTAAg	CTaATACTAA	C.....ACca	tgcAatGCTT
Hapo1234	tAgcCTGATC	CAacATTAAg	CTgATACTAA	CaaacaACgt	gtaAtgGCTT
Consensus	-A--CTGATC	CA--ATTAAg	CT-ATACTAA	C-----AC--	---A--GCTT
	501				550
Mgen1073	CATTAAcAAG	GaTTTGCTTC	TTgCTaGAAA	tgGGT..AAA	AaCggACTgt
Hapo1234	CATTAAaAAG	GcTTTGCTTC	TTcCTgGAAA	ctGGTgaAAA	AtCaaACctT
Consensus	CATTAA-AAG	G-TTTGCTTC	TT-CT-GAAA	--GGT--AAA	A-C--AC--T
	551				600
Mgen1073	GgTcTGTAtA	CcTTCaATGC	AGCTTaTGTG	TTGTCTTttC	C..tgAAatG
Hapo1234	GtTgTGTAcA	CCcTCgATGC	AGCTTcTGTG	TTGTCTTcaC	CcagaAAatgG
Consensus	G-T-TGTA-A	CC-TC-ATGC	AGCTT-TGTG	TTGTCTT--C	C----AA--G

Figure 11

	601				650
Mgen1073	GtAATGAcTc	CCaAtAgtGg	cAAccAgggG	tacaATaCT.	.....TGCA
Hapo1234	GgAATGAtTt	CCcAaAtgGc	aAAgaAacaG	agtgATgCTa	tctatcTGCA
Consensus	G-AATGA-T-	CC-A-A--G-	-AA--A---G	----AT-CT-	-----TGCA
	651			exon79	700
Mgen1073	CacTTTGTA	A....cTCTT	TCTTTCTCTT	TGTTTTCCAG	GACACAATGT
Hapo1234	CctTTTGTA	AgtctgTCTT	TCTTTCTCTT	TGTTTTCCAG	GACACAATGT
Consensus	C--TTTGTA	A-----TCTT	TCTTTCTCTT	TGTTTTCCAG	GACACAATGT
	701				750
Mgen1073	AGGAAGcCTT	TTCCACATGG	CAGATGATTT	GGGCAGAGCG	ATGGAGTCCT
Hapo1234	AGGAAGtCTT	TTCCACATGG	CAGATGATTT	GGGCAGAGCG	ATGGAGTCCT
Consensus	AGGAAG-CTT	TTCCACATGG	CAGATGATTT	GGGCAGAGCG	ATGGAGTCCT
	751				800
Mgen1073	TAGTtTCAGT	CATGACAGAT	GAAGAAGGAG	CAGAATAAAT	GTTTTCACAC
Hapo1234	TAGTaTCAGT	CATGACAGAT	GAAGAAGGAG	CAGAATAAAT	GTTTTCACAC
Consensus	TAGT-TCAGT	CATGACAGAT	GAAGAAGGAG	CAGAATAAAT	GTTTTCACAC
	801				850
Mgen1073	TCCTGATTCC	CGCATGGTTT	TTATAATATT	CgTACAACAA	AGAGGATTAG
Hapo1234	TCCTGATTCC	CGCATGGTTT	TTATAATATT	CaTACAACAA	AGAGGATTAG
Consensus	TCCTGATTCC	CGCATGGTTT	TTATAATATT	C-TACAACAA	AGAGGATTAG
	851				900
Mgen1073	ACAGTAAGAG	TTTACAAGAA	ATaAAATCTA	TATTTTTGTG	AAGGGTAGTG
Hapo1234	ACAGTAAGAG	TTTACAAGAA	AT.AAATCTA	TATTTTTGTG	AAGGGTAGTG
Consensus	ACAGTAAGAG	TTTACAAGAA	AT-AAATCTA	TATTTTTGTG	AAGGGTAGTG
	901				950
Mgen1073	GTAcTATACT	GTAAGTTTCA	GTAAGTTTCTA	AGTCTGTTAT	TGTTTTGTTA
Hapo1234	GTAtTATACT	GTAAGTTTCA	GTAAGTTTCTA	AGTCTGTTAT	TGTTTTGTTA
Consensus	GTA-TATACT	GTAAGTTTCA	GTAAGTTTCTA	AGTCTGTTAT	TGTTTTGTTA
	951				1000
Mgen1073	ACAATGGCAG	GTTTTACACG	TCTATGCAAT	TGTACAAAAA	AGTTAaAAGA
Hapo1234	ACAATGGCAG	GTTTTACACG	TCTATGCAAT	TGTACAAAAA	AGTTAtAAGA
Consensus	ACAATGGCAG	GTTTTACACG	TCTATGCAAT	TGTACAAAAA	AGTTA-AAGA
	1001				1050
Mgen1073	AA...ACATG	TAAAATCTTG	ATAGCTAAAT	AACTTGCCAT	TTCTTTATAT
Hapo1234	AAactACATG	TAAAATCTTG	ATAGCTAAAT	AACTTGCCAT	TTCTTTATAT
Consensus	AA---ACATG	TAAAATCTTG	ATAGCTAAAT	AACTTGCCAT	TTCTTTATAT
				begin inversion@1100	
	1051				1100
Mgen1073	GGAACGCATT	TTGGGTGTG	TAAAAATTTA	TAACAGTTAT	AAAGAAAGAt
Hapo1234	GGAACGCATT	TTGGGTGTG	TAAAAATTTA	TAACAGTTAT	AAAGAAAGAA
Consensus	GGAACGCATT	TTGGGTGTG	TAAAAATTTA	TAACAGTTAT	AAAGAAAGA-
	1101				1150
Mgen1073	TgtAAActaA	Agtgtgcttt	AtAAAAaAAg	ttgtTtataA	AaaccctAa
Hapo1234	TtatAAaggA	A.....aa	AgAAAAAtAa	gcaaTggacA	AgtggtgaAg
Consensus	T---AA---A	A-----	A-AAAA-AA-	----T----A	A-----A-
	1151				1200
Mgen1073	acaaacACaC	AcGcacaCAC	AcacAcacac	AcacaCaCaC	AcaCaCaCTG
Hapo1234	ctgtgaACTc	AgGtgtgCAC	AattAtcagg	AacacCcCAa	AacCAaAgTG
Consensus	-----AC-C	A-G---CAC	A---A-----	A----C-CA-	A--CA-A-TG
	1201				1243
Mgen1073	AGGcAGcAca	ttgtTttGcA	ttacTtTagc	gTGTatcaTA	t..
Hapo1234	AGGtAGaAat	agcaTgaGaA	gccgTgTttg	aTGTaatTA	att
Consensus	AGG-AG-A--	----T--G-A	----T-T---	-TGT-----TA	---

Figure 11 (cont'd)



0906264.092801

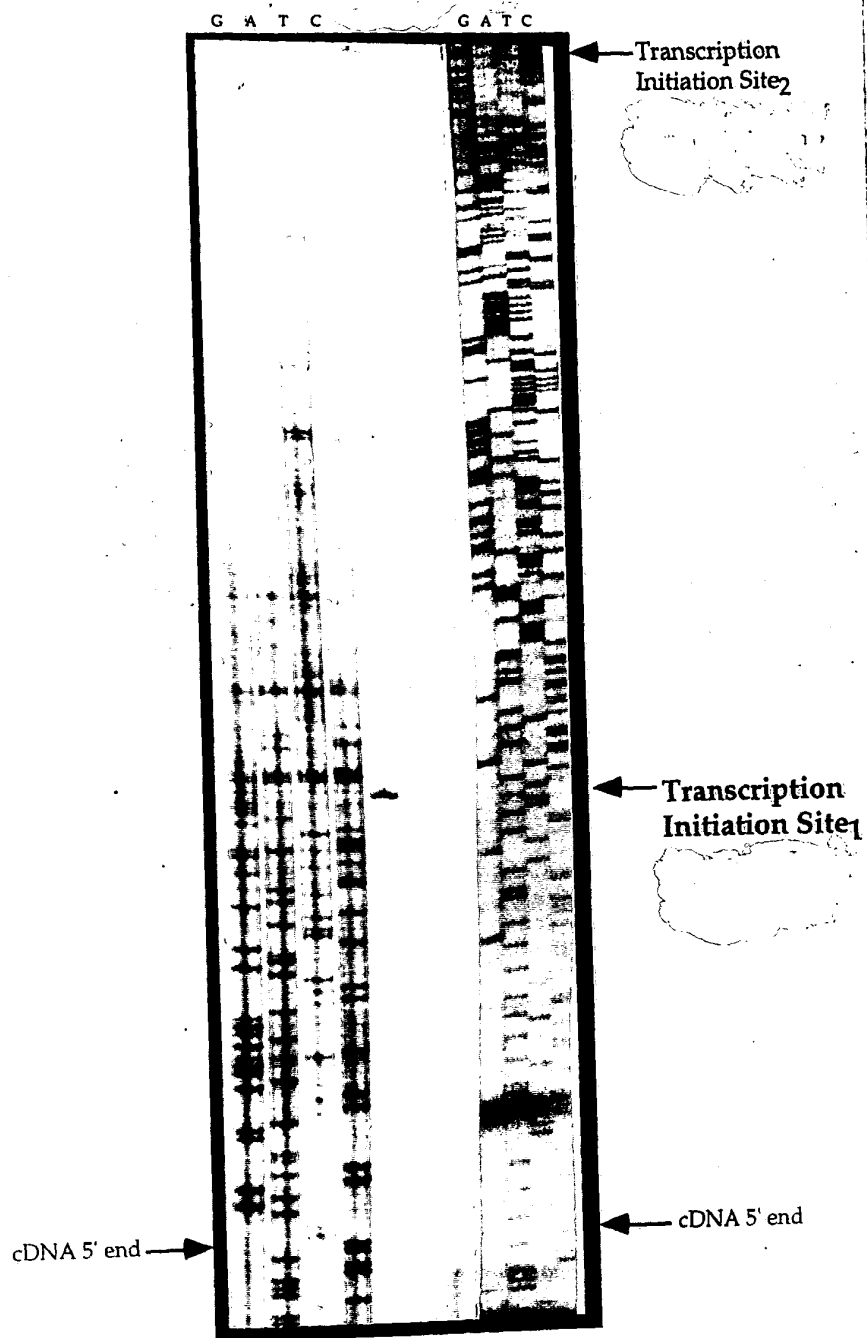


Figure 12A

-70 bp from 5' end of apo-4

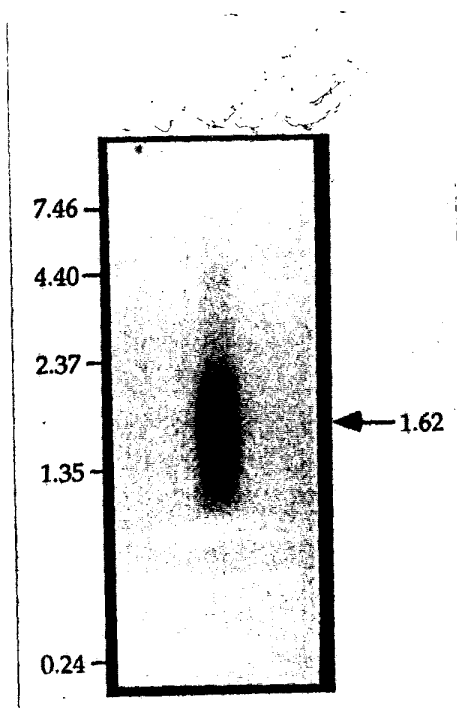
Inr = GCCC TCAT TCTG GAGAC

apo-4 = GCGG TGAT GGTG GCAGT - 48% perfect homology with Inr

71% match on type of base

(purine vs. pyrimidine)

**Figure 12B**



**Figure 13**

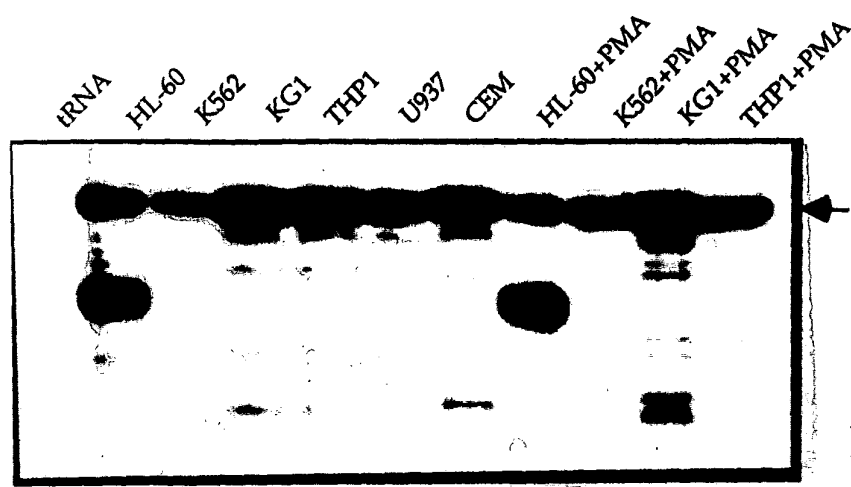
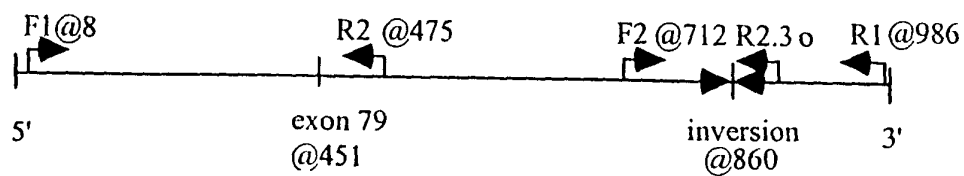


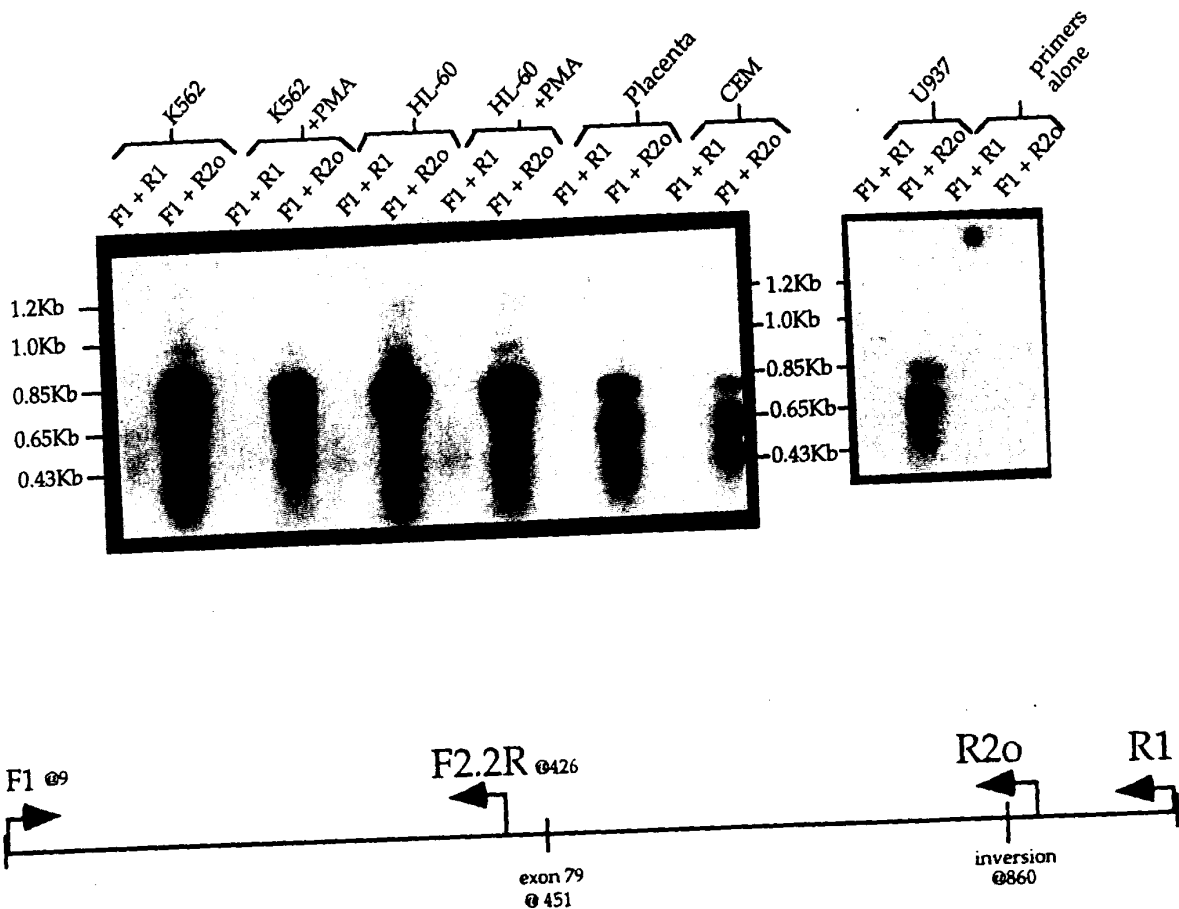
Figure 14



**Figure 15**

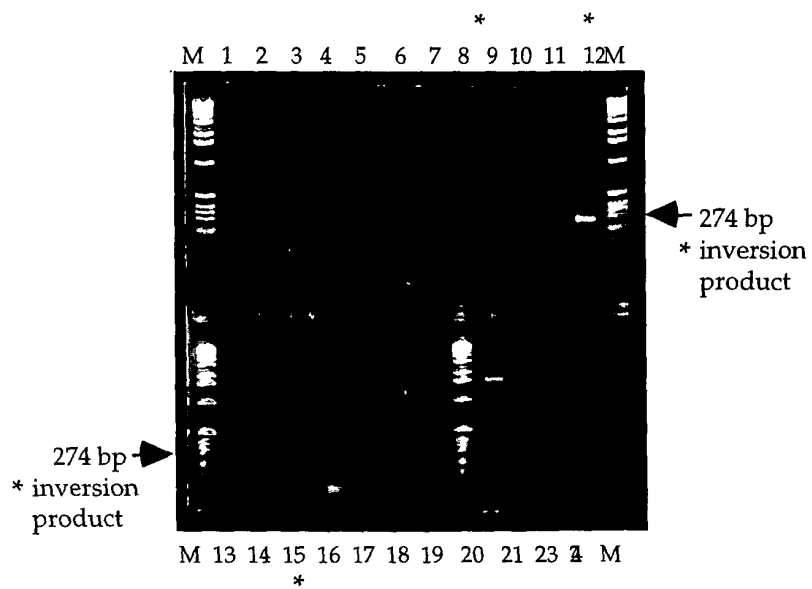
096664-096660

095664-1329960



The Apo-dystrophin cDNA

Figure 16



**Figure 17A**

T03250-19299660

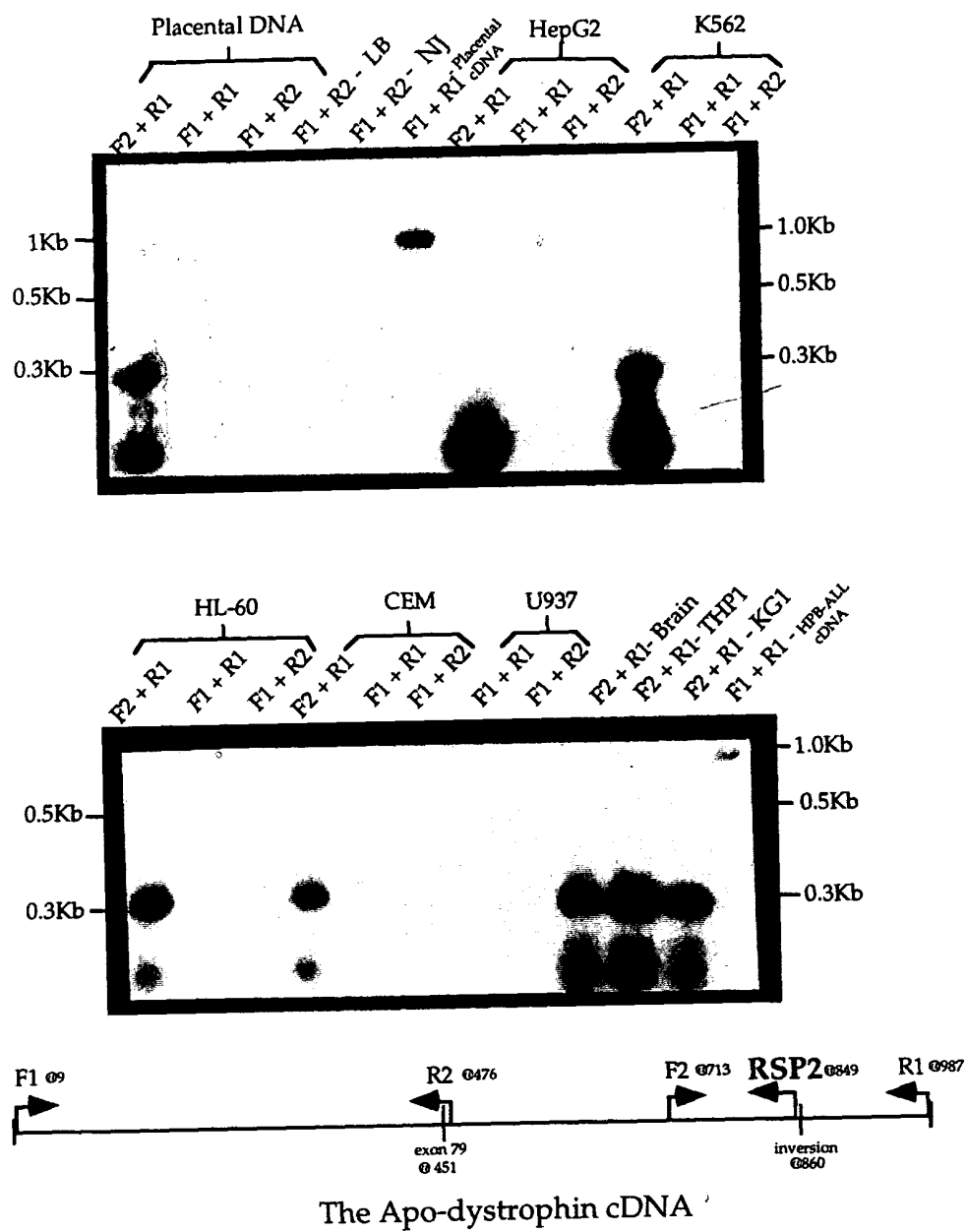
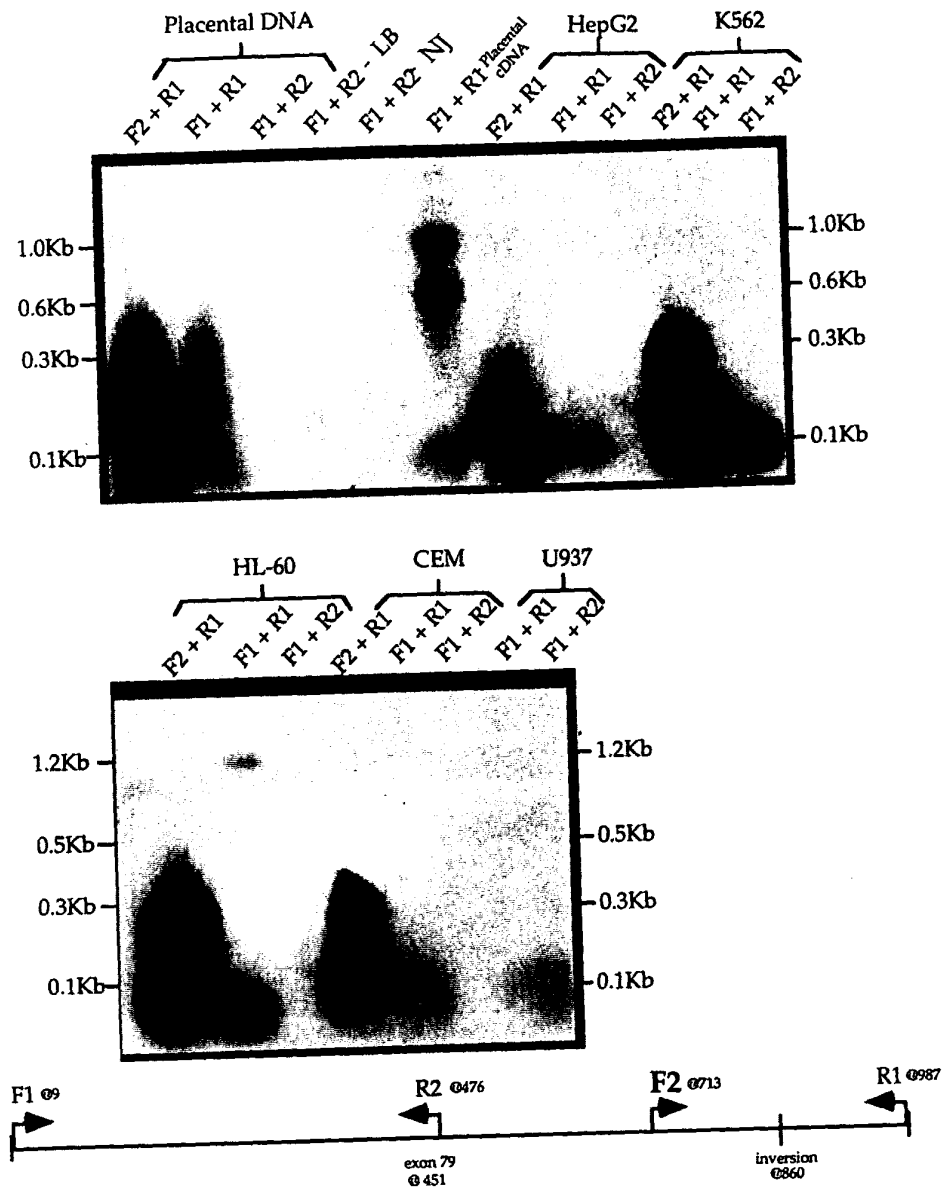


Figure 17B

0956364-09501



The Apo-dystrophin cDNA

Figure 17C



A.

12/23bp spacer

CACAGTG-----	-----ACAAAAACC
heptamer	nonamer

**Figure 18A**

B.

	inversion breakpoint <sub>1</sub>					
	11640	11650	11660		11670	11680
	*	*	*		*	*
dystrophin	T	TTATAACAGT	TATAAAGAAA	GA^TTGTAAAC	<b>TAAAGTGTGC</b>	
	A	AATATTGTCA	ATATTTCTTT	CT^AACATTG	ATTCACACG	
				a		
apo-4 cDNA	840	850			870	
[ 138 ]	T	TTATAACAGT	TATAAAGAAA	GA^TTaTAAAg	gAAAAaGaaa>	
	^	^^^^^^^^^^	^^^^^^^^^^	^^ ^v^v^v^v	v^^^vv^vvv	
dystrophin	T	TTATAACAGT	TATAAAGAAA	GA^TTGTAAAC	TAAAGTGTGC	

---

	11690	11700	11710	11720	11730
	*	*	*	*	*
dystrophin	TTTATAAAAA	AAAGTTGTTT	<b>ATAAAAACCC</b>	<b>CTAAAAACAA</b>	AACAAACACA
	AAATATTTTT	TTTCAACAAA	TATTTTGGG	GATTTTGTGTT	TTGTTTGTGT
apo-4 cDNA	880	890	900	910	920
[ 138 ]	aTaAaAtggA	cAAGTgGTga	ATgtgAACTC	aggtgtgCAc	AAttAtCAgg>
	v^v^v^vvv^	v^^^v^^v	^^vvv^^^v^	vvvvvvv^^v	^^vv^v^^vv
dystrophin	TTTATAAAAA	AAAGTTGTTT	ATAAAAACCC	CTAAAAACAA	AACAAACACA

---

	11740	11750
	*	*
dystrophin	CACACACACA	<b>CATACACACA</b>
	GTGTGTGTGT	GTATGTGTGT
apo-4 cDNA	940	950
[ 138 ]	aACAC-CcCA	-AaAC-CAaA>
	v^^^^ ^v^^	^v^^ ^^v^
dystrophin	CACACACACA	CATACACACA

**Figure 18B**

T09220-4929966

09966264-092801

	13130	13140	13150	13160	13170
	*	*	*	*	*
dystrophin	AATTAGCTTT	TGGAGAGTGG	GTTTTGTCCA	TTATTAATAA	TTAATTAATT
	TTAATCGAAA	ACCTCTCACC	CAAAACAGGT	AATAATTATT	AATTAATTAA

inv rsion breakpoint2

	13180	13190	13200	13210	13220
	*	*	*	*	*
dystrophin	AACATCAAAC	ACGGCTTCTC	ATGCTATTTT	TACCTCACTT	TGGTTTTGGG
	TGTAGTTTGG	TGCCGAAGAG	TACGATAAAG	ATGGAGTGAA	ACCAAAACCC

	980	970	960	950	940
apo-4	<AACATCAAAC	ACGGCTTCTC	ATGCTATTTT	TACCTCACTT	TGGTTTTGGG
	^^^^^^^^	^^^^^^^^	^^^^^^^^	^^^^^^^^	^^^^^^^^
dystrophin	AACATCAAAC	ACGGCTTCTC	ATGCTATTTT	TACCTCACTT	TGGTTTTGGG

	13230	13240	13250	13260	13270
	*	*	*	*	*
dystrophin	GTGTTCTCTGA	TAATTGTGCA	CACCTGAGTT	CACAGCTTCA	CCACTTGTCC
	CACAAGGACT	ATTAACACGT	GTGGACTCAA	GTGTCGAAGT	GGTGAACAGG

	930	920	910	900	890
apo-4	<GTGTTCTCTGA	TAATTGTGCA	CACCTGAGTT	CACAGCTTCA	CCACTTGTCC
	^^^^^^^^	^^^^^^^^	^^^^^^^^	^^^^^^^^	^^^^^^^^
dystrophin	GTGTTCTCTGA	TAATTGTGCA	CACCTGAGTT	CACAGCTTCA	CCACTTGTCC

	13280	13290	13300	13310	13320
	*	*	*	*	*
dystrophin	ATTGCGTTAT	TTTCTTTTTC	CTTTATAAAT	CTTCTTTTTT	CCTTCATAAT
	TAACGCAATA	AAAGAAAAAG	GAAATATTAA	GAAAGAAAAA	GGAAGTATTA

|

inversion breakpoint3

	880	870	860	850	840
apo-4	<ATTGCGTTAT	TTTCTTTTTC	CTTTATAAAT	CTTCTTTTTa	aacTgtTATa
	^^^^^^^^	^^^^^^^^	^^^^^^^^	^^^^^^^^	^^^^^^^^
dystrophin	ATTGCGTTAT	TTTCTTTTTC	CTTTATAAAT	CTTCTTTTTT	CCTTCATAAT

Figure 18C

5'

11654

AAAGAAAGAT

|||||

TTCTTTCTT

13299

3'

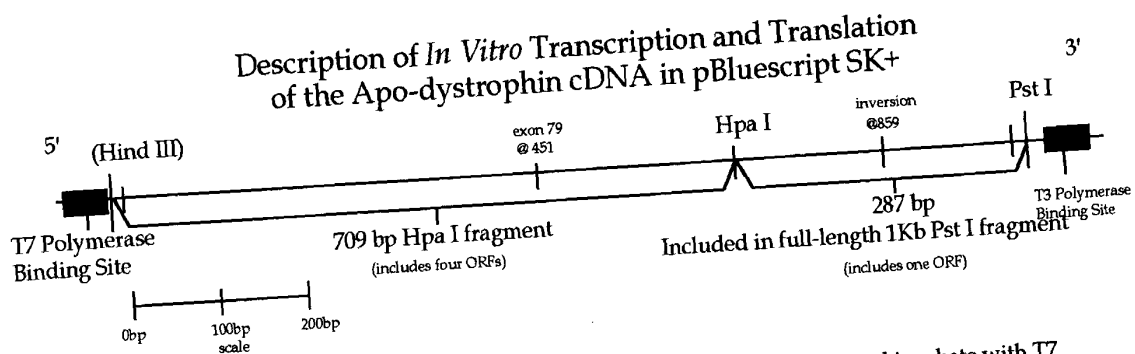
inversion breakpoint

**Figure 18D**

inversion @ 860

841 TAACAGTTATAAGAAAGAATTATAAAGGAAAAAGAAAATAACGCAATGGACAAGTGGTG 900  
-----+-----+-----+-----+-----+  
ATTGTCAATAATTTCTTCTTAATATTTCTTTCTTTATTGCGTTACCTGTTCACCAC

### Figure 19

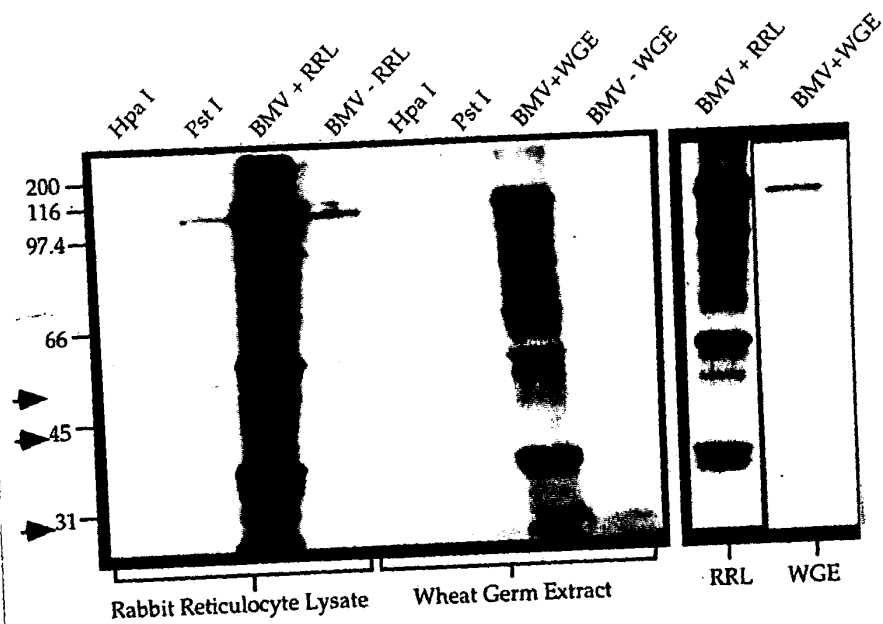


Linearize plasmid with either Hpa I (truncated) or Pst I (full length). Gene Clean and incubate with T7 polymerase and dNTPs to produce RNA *in vitro*.

Incubate RNA with Wheat Germ Extract or Rabbit Reticulocyte Lysates to produce *in vitro* translation

Separate translation products by SDS-PAGE. Fix, Amplify and Dry Gel. Perform Autoradiography

**Figure 20**



**Figure 20A**

**Figure 20B**

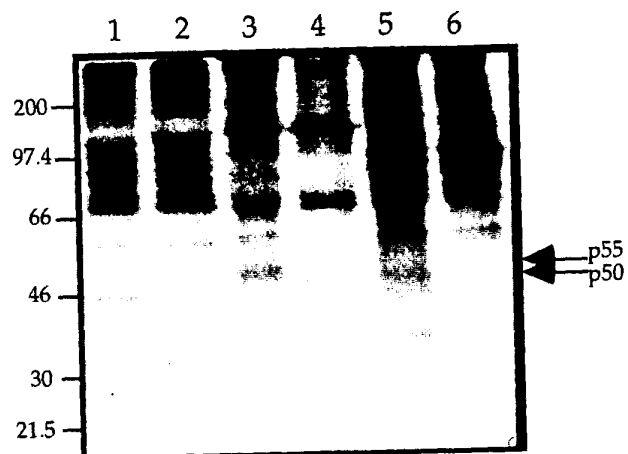


Figure 21

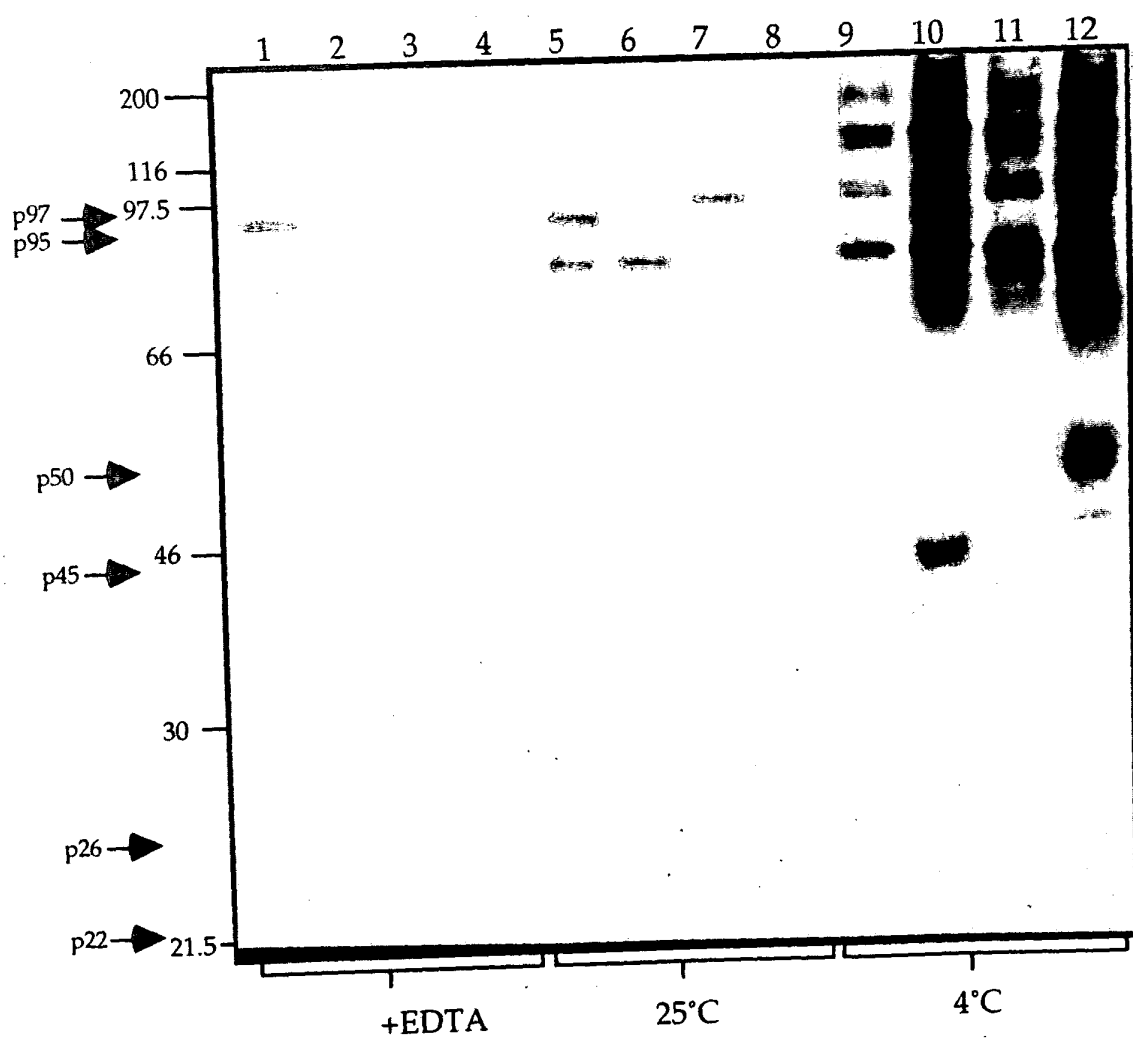


Figure 22

FD3260-49299660

09664-19660

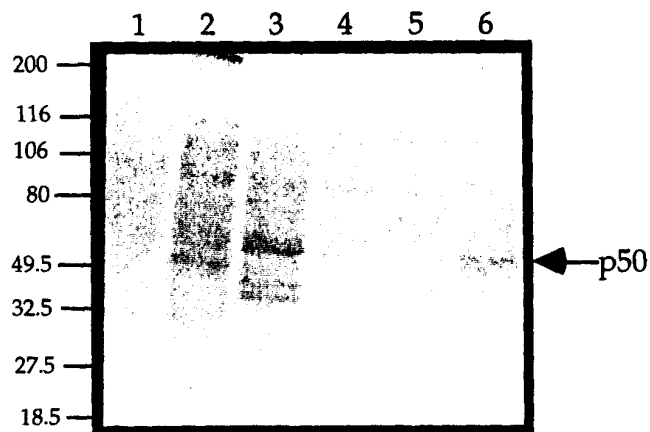


Figure 23

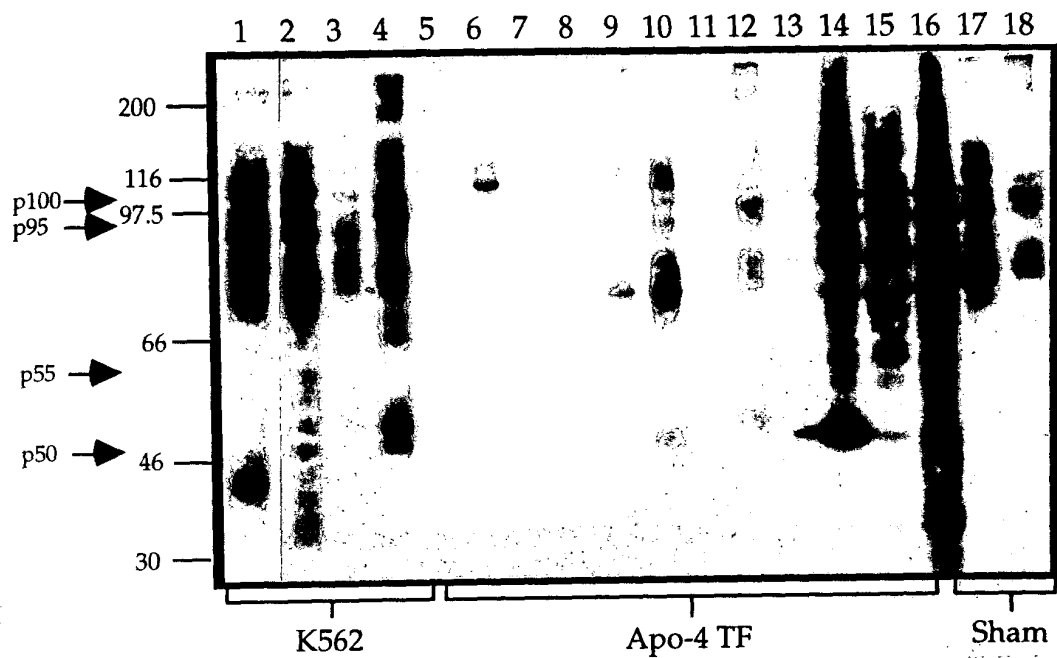


Figure 24

09956264-092801

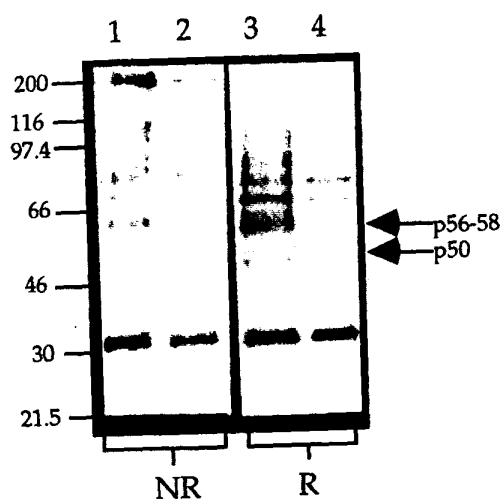


Figure 25A



Figure 25B

H2 starting at second methionine - 321 bp, predicted weight = 17.4Kd + 1 N-glycosylation site + 20.4 Kd.

**Figure 26A**

Splice sites for peptide

MYPIMEYSCSDRNLVLIYGILLIYIYIGNLNMKKEQNKCFTTPDSRMVFIIIFIQQRGLDSKSLQEINL  
YFCEGFYTSMQLYKKVIRKLHKITQWTRTPQNQSEVEIA 107

**Figure 26B**

Start	Exon No.	Exon Position	Exon Length	Intron No.	Intron Position	Intron Length
@88 bp	78.3	@74-180	106 bp	79.1	@181-529	349 bp
	79.1	@530-654	125 bp	79.4	@655-720	66 bp
	79.4	@721-769	49 bp	79.55	@770-875	105 bp
	79.55	@876-893	18 bp	79.75	@894-932	39 bp
	79.85	@933-966	33 bp			

Hydrophobicity Scale KD; Candidate membrane-spanning segments:

Certain 1 12- 32 1.8833

**Figure 26C**

09066264-092801



0996664-092601

Predicted TM structure

- > : Too long to be significant
- < : Too short to be significant
- LI : Loop length
- KR : Number of Lys and Arg

- KR Diff : Positive charge difference
- CE : Net charge energy
- CE Diff : Net charge difference
- CH Diff : Charge difference over N-term segments

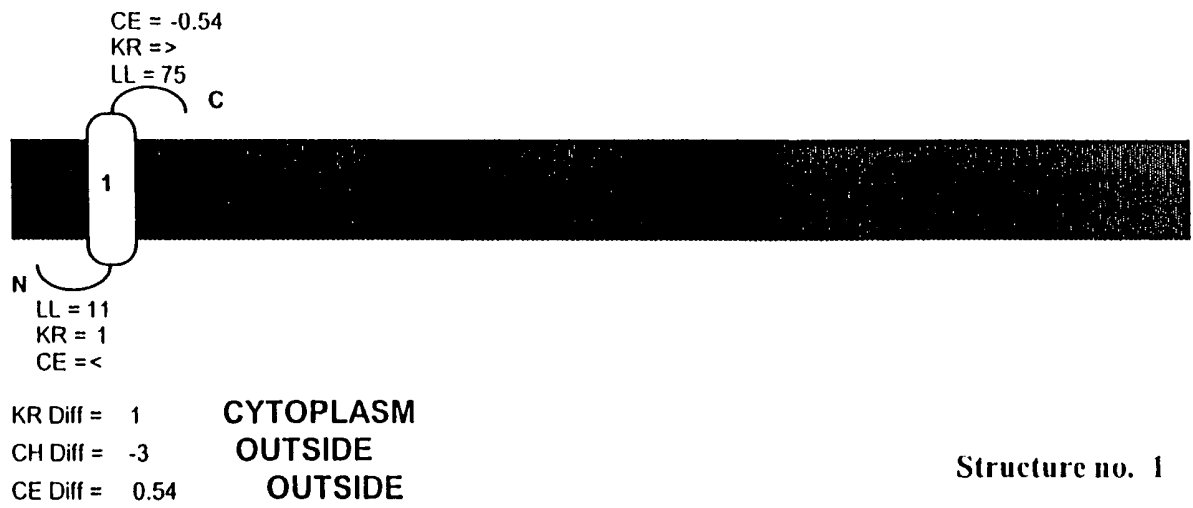


Figure 26D

Variable	Mean	SD	Min	Max
Age	38.5	12.5	18	65
Gender	Male	Female		
Marital status	Married	Single		
Education	High school	College		
Occupation	Manager	Worker		
Income	\$30,000	\$40,000		
Health status	Good	Fair		
Exercise frequency	Weekly	Monthly		
Stress level	Low	High		
Sleep quality	Good	Poor		
Dietary habits	Healthy	Unhealthy		
Alcohol consumption	None	Occasional		
Tobacco use	Non-smoker	Smoker		
Family size	2	3		
Work hours	40	50		
Commuting time	30	45		
Home ownership	Owner	Renter		
Neighborhood safety	Safe	Unsafe		
Access to green spaces	Yes	No		
Proximity to public transport	Close	Far		
Local amenities	Many	Few		
Community involvement	Active	Passive		
Perceived quality of life	High	Low		

### Figure 27A

Certain	1	41-61	1.9073
Putative	2	101-121	0.8052
Certain	3	132-152	1.2552
Putative	4	217-237	1.1833
Putative	5	254-274	0.9240

### Figure 27B

> : Too long to be significative  
 < : Too short to be significative  
 LI : Loop length  
 KR : Number of Lys and Arg

KR Diff : Positive charge difference  
 CE : Net charge energy  
 CE Diff : Net charge difference  
 CH Diff : Charge difference over N-term segments

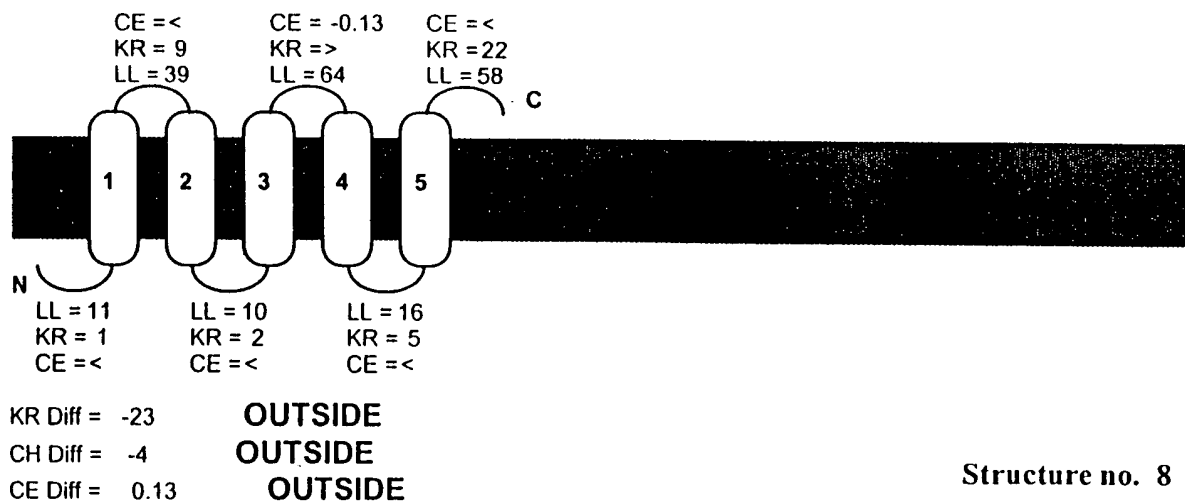


Figure 27C

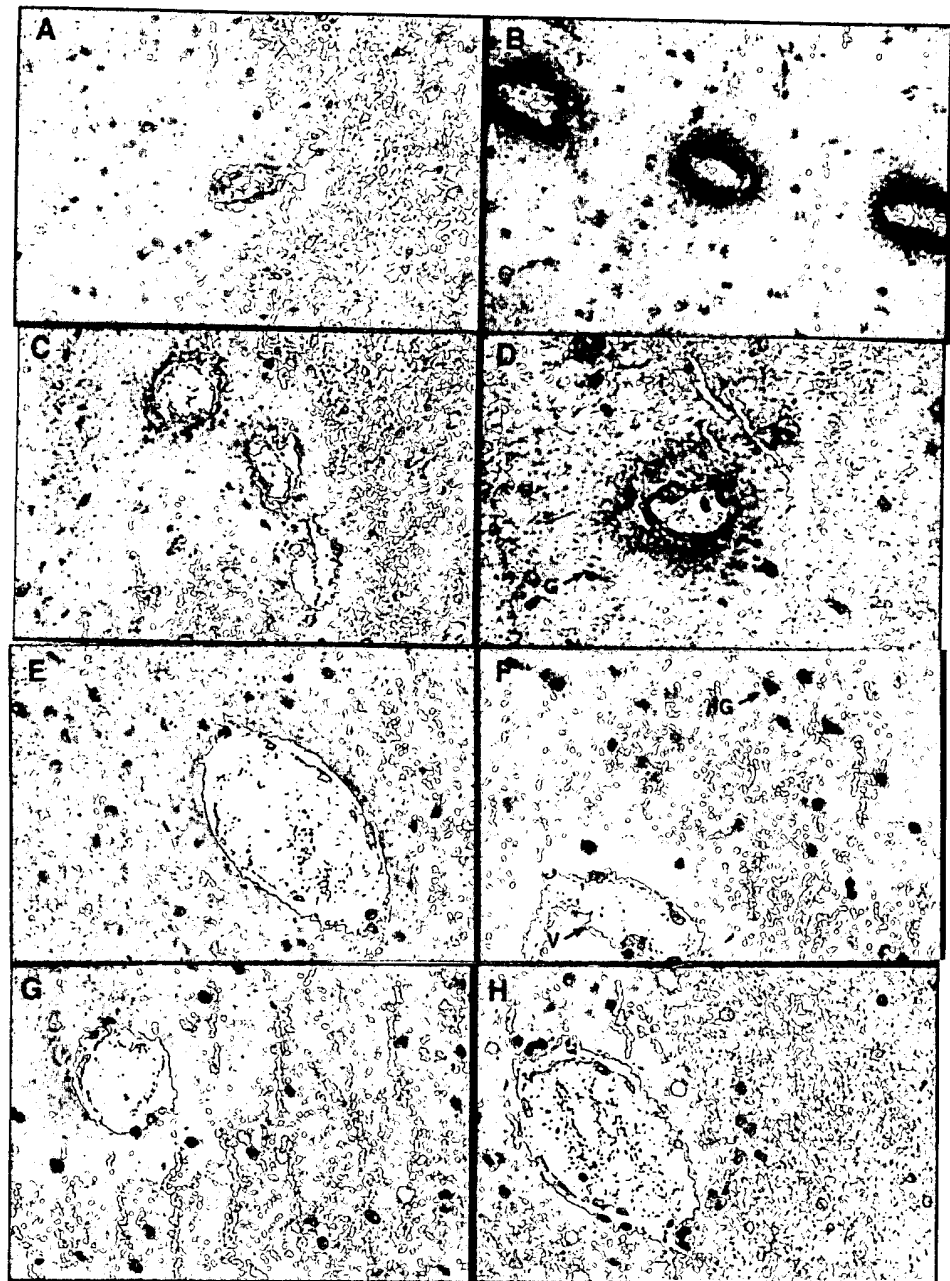
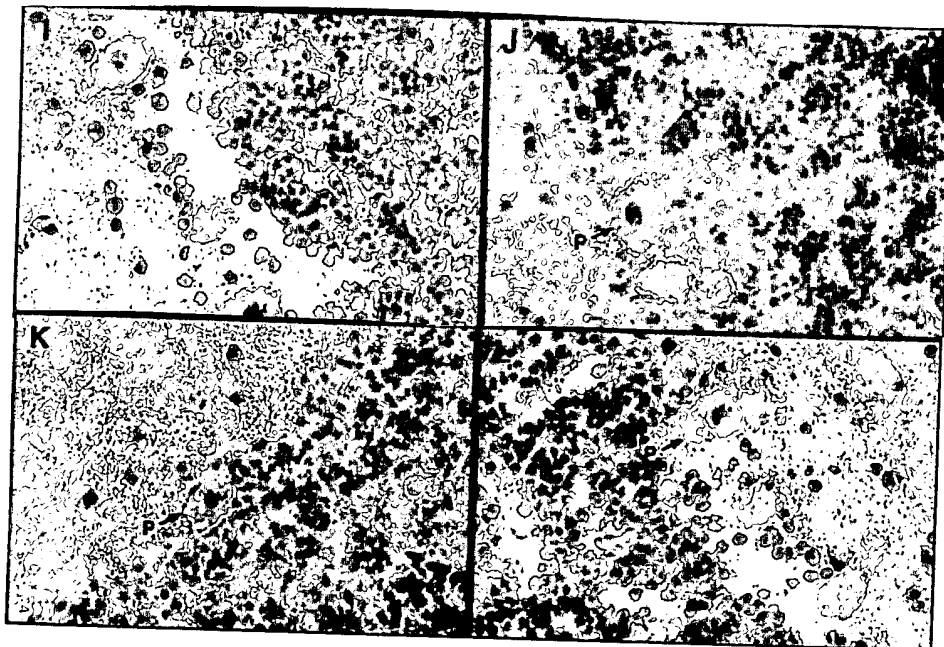


Figure 28

0966264-092801



**Figure 28 (cont'd)**

108260-4329960

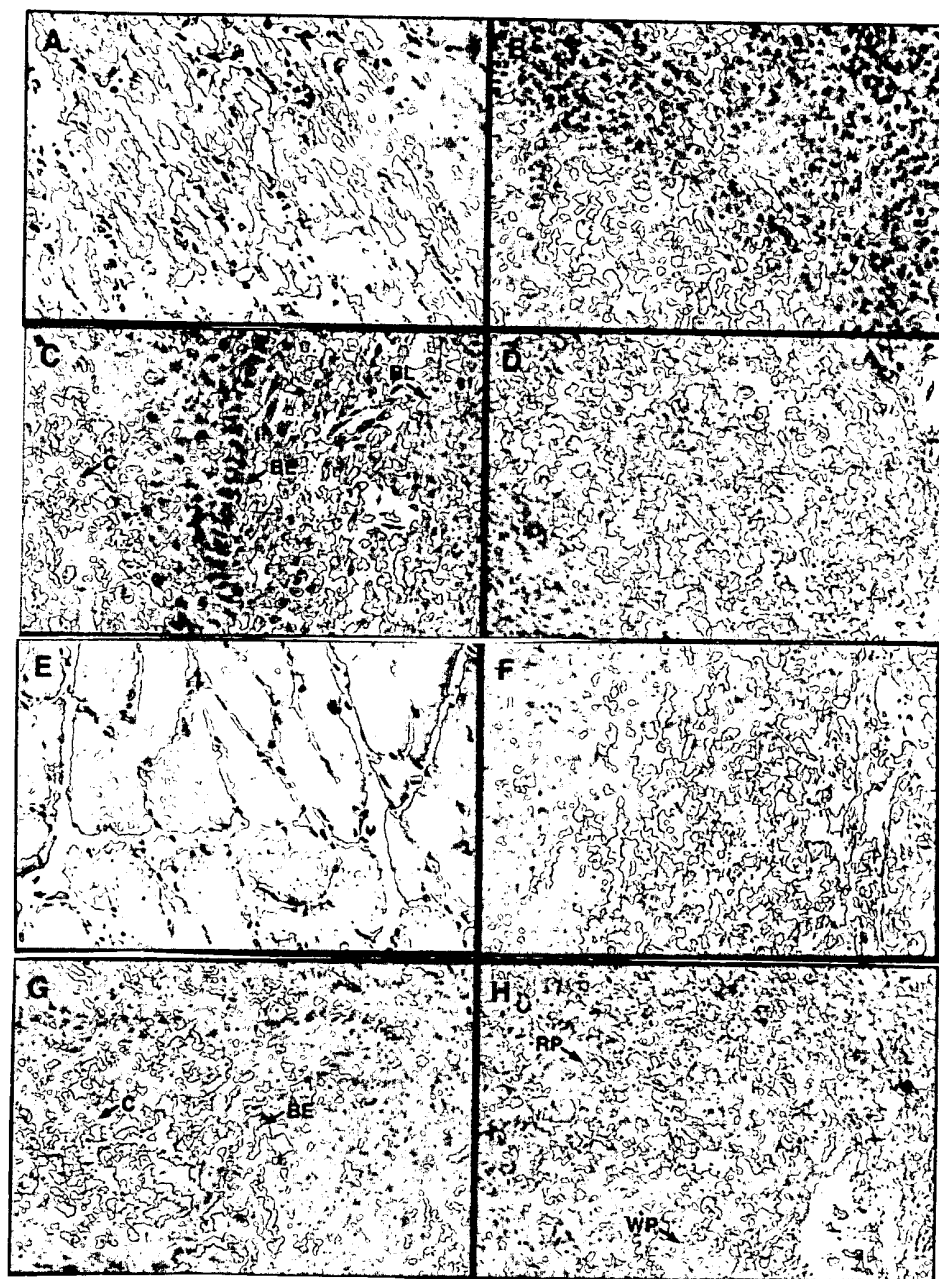
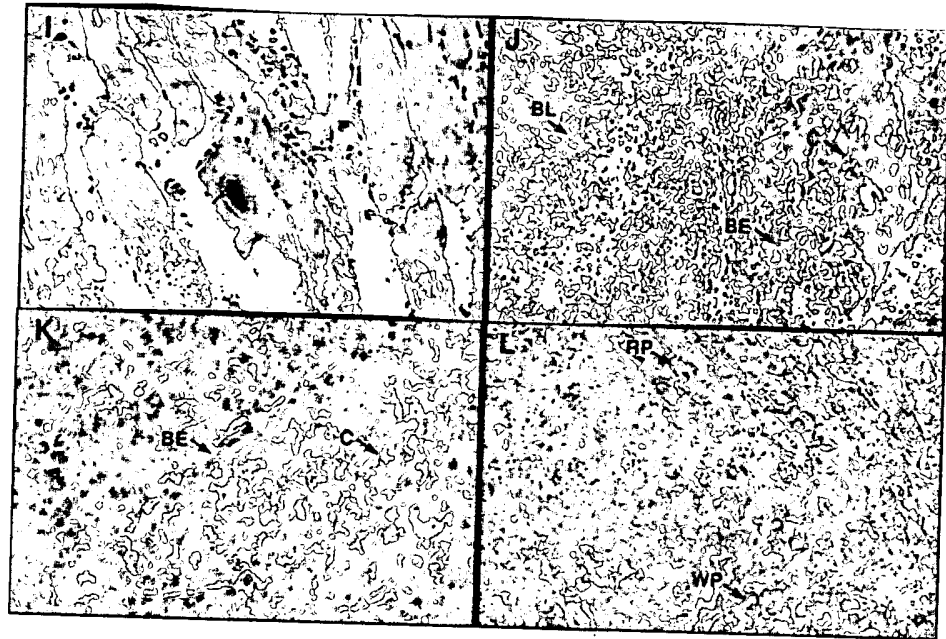


Figure 29



**Figure 29 (cont'd)**

0996234-09201

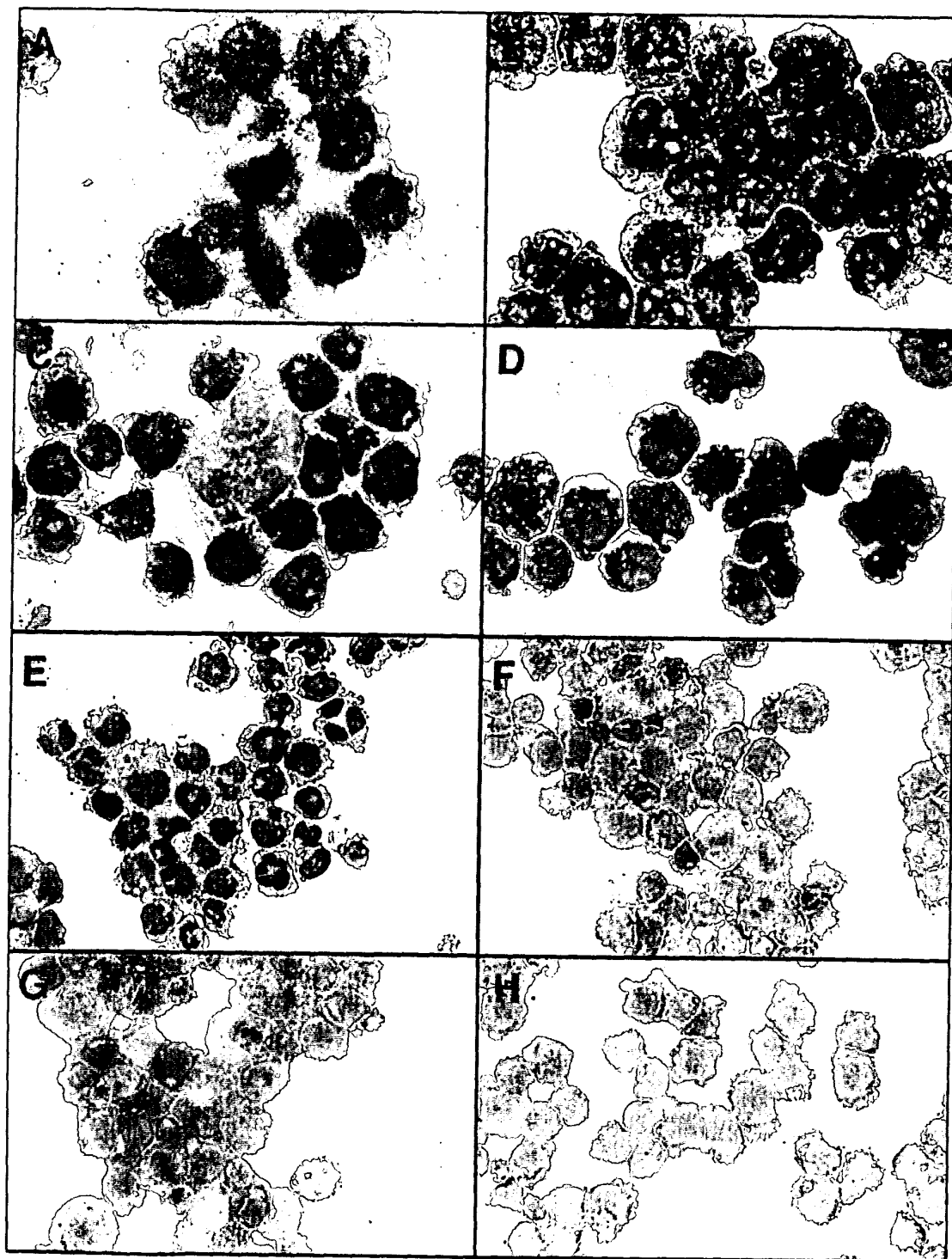


Figure 30



00966264-092801

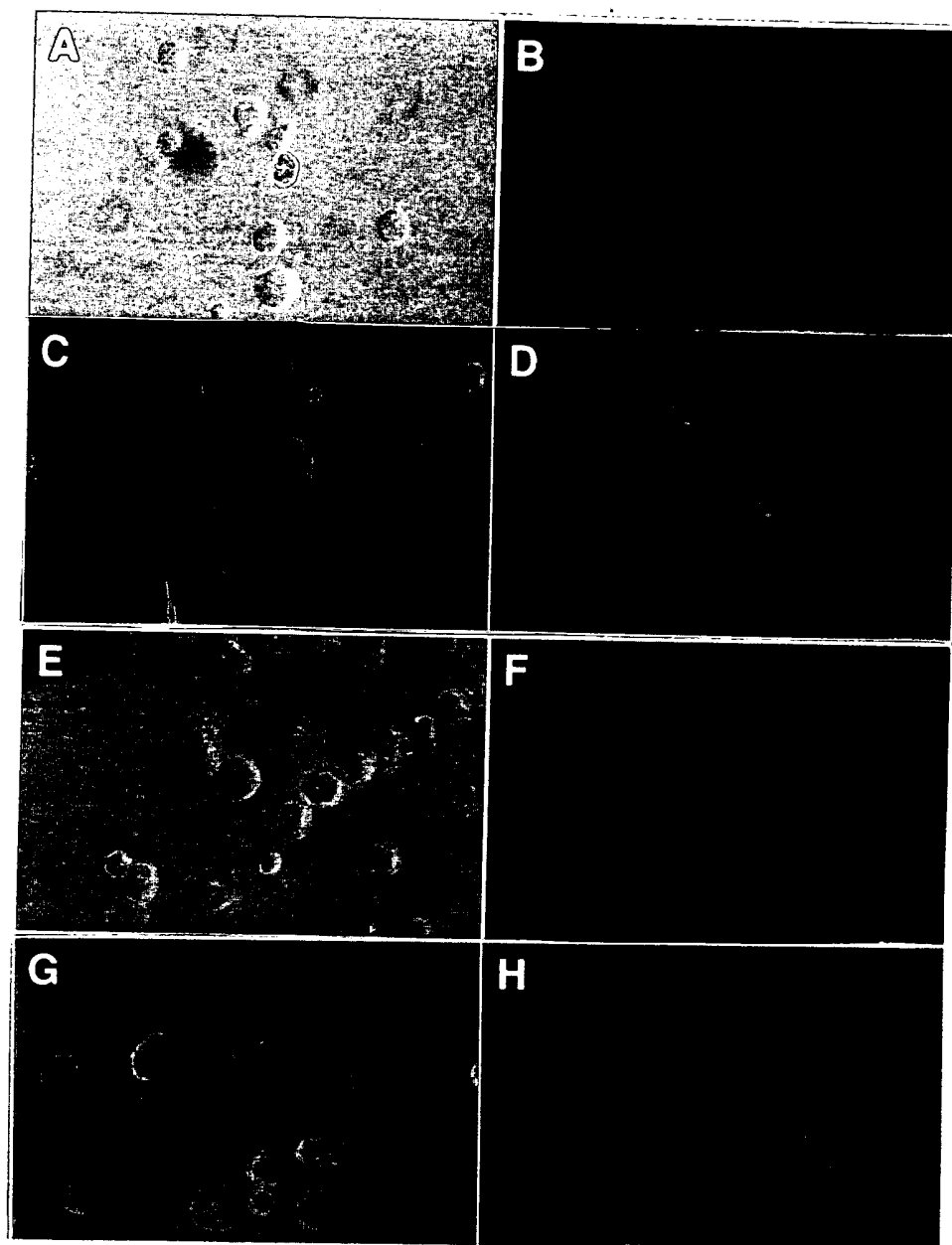
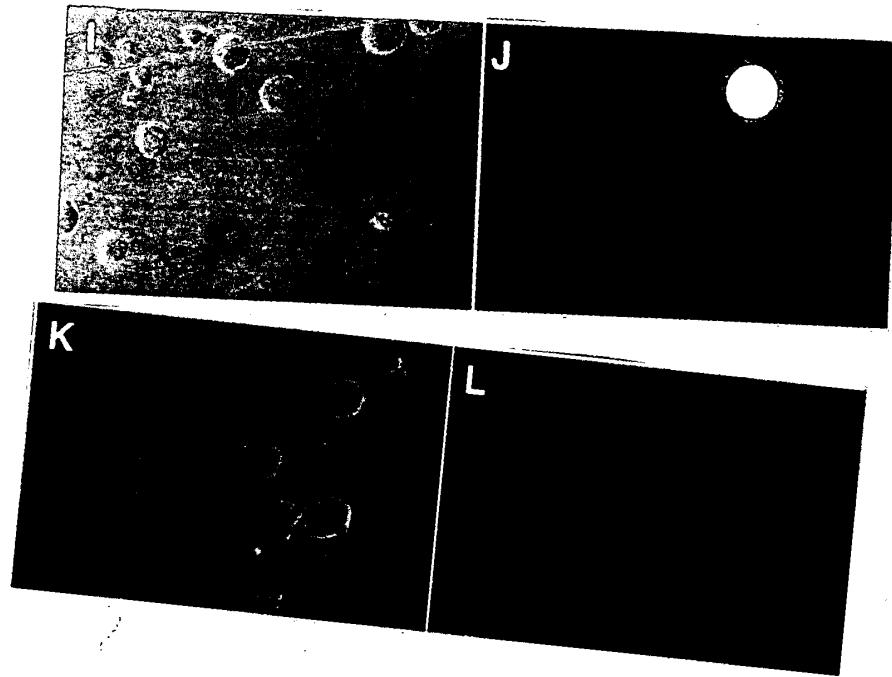


Figure 31

108260-49299660



**Figure 31 (cont'd)**

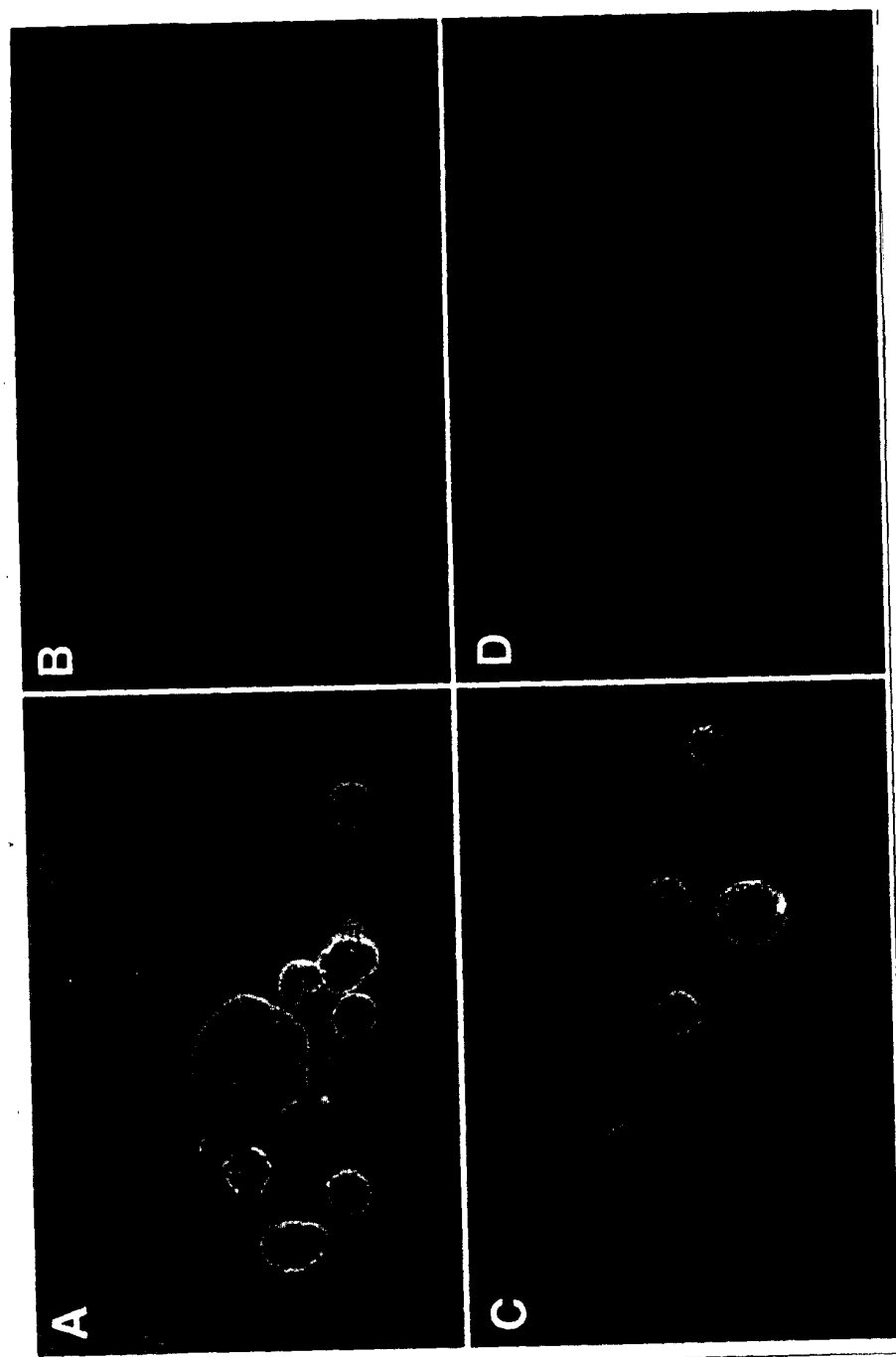


Figure 32

099664-19299660

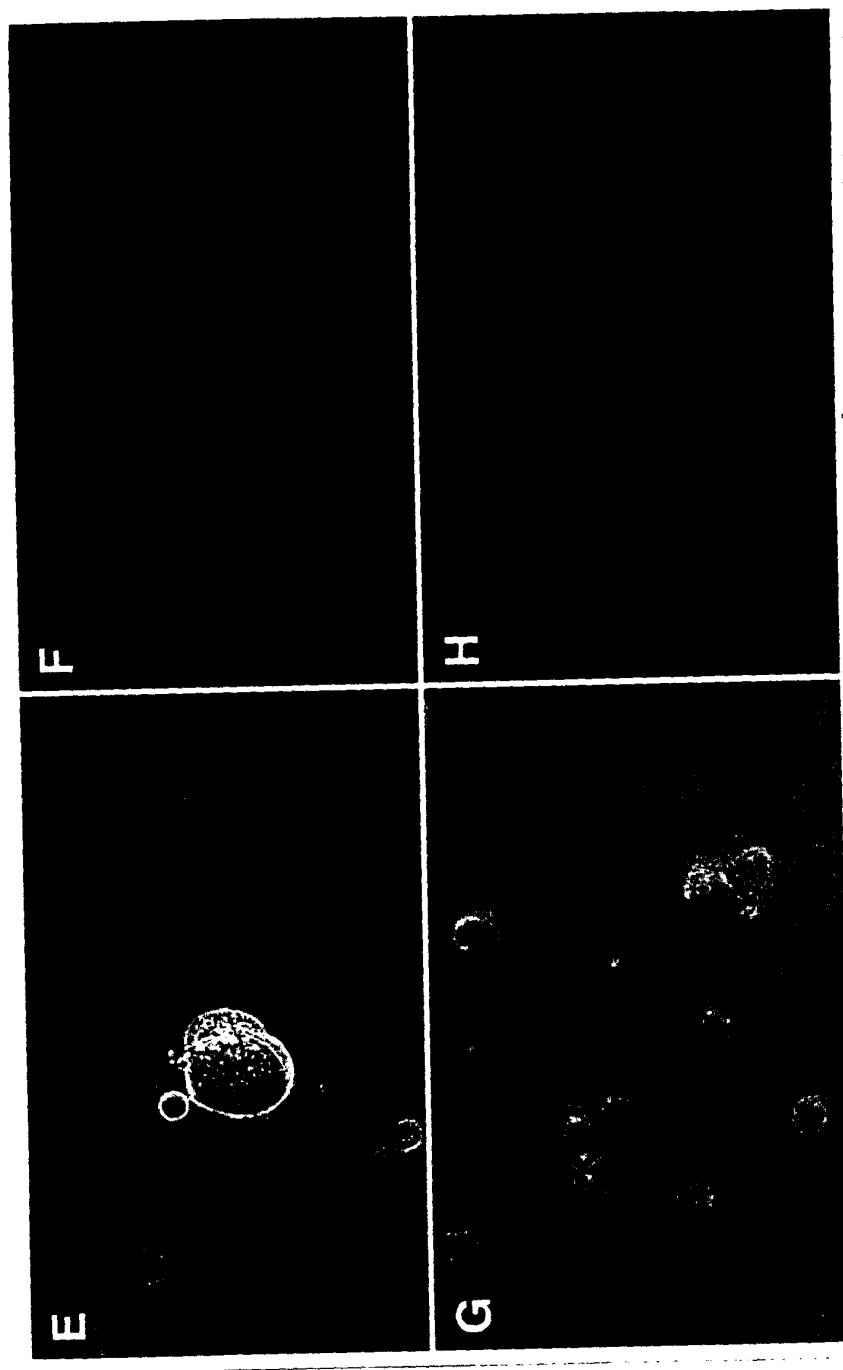


Figure 32 (cont'd)

FO3260-49299660

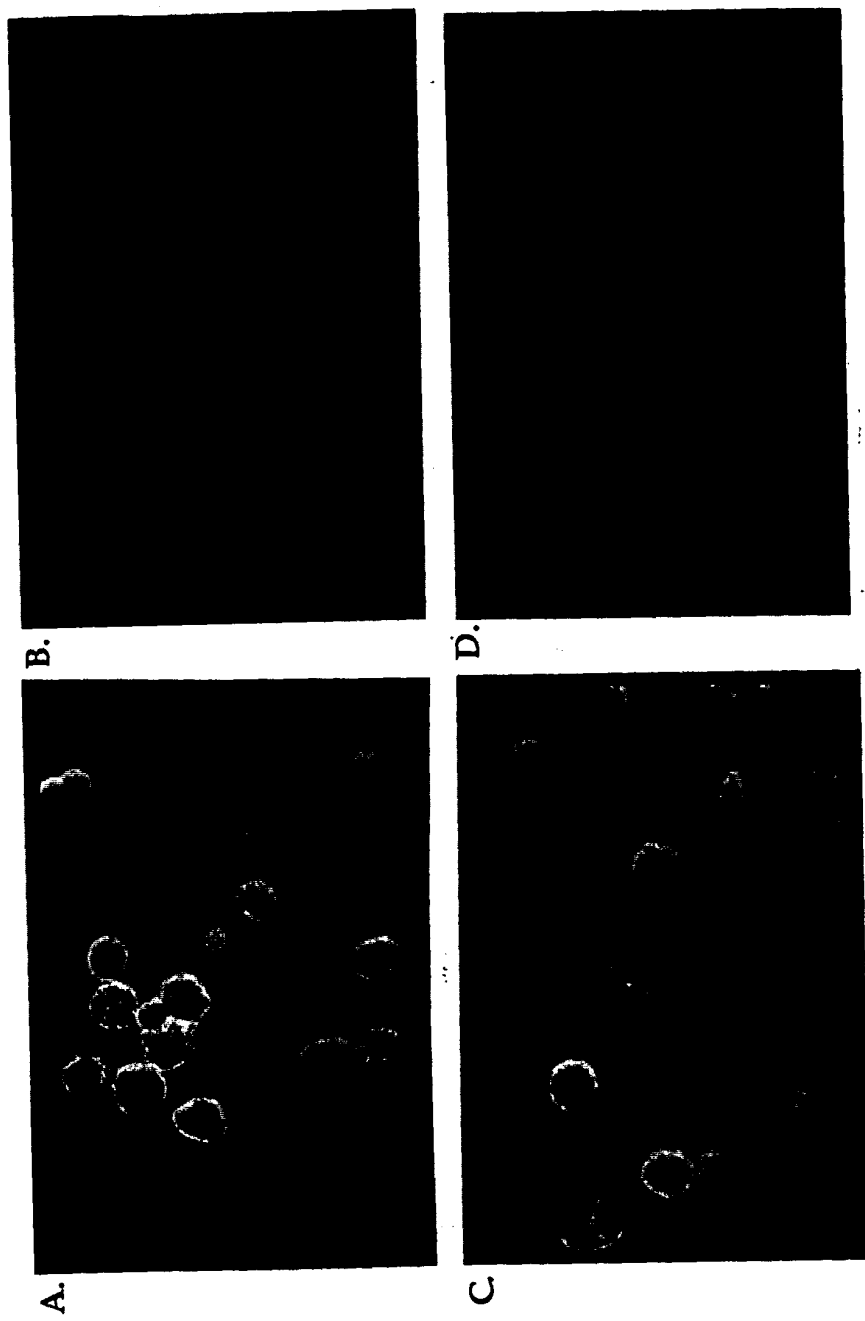


Figure 33

**Additional Oligonucleotide primers used for apo-dystrophin-4  
southern blotting and sequencing**

**FORWARD**

GTT CGT TAA TAC AAG TAG	F2.3(@28)	(SEQ ID NO 15)
GCC AAG GTG GAA AAG ATG	F2.2(@73)	(SEQ ID NO 16)
CCA GTA GCC TGA TCC AAC	F3.2(@208)	(SEQ ID NO 17)
GGC TTC ATT AAT AAG	F3.1(@257)	(SEQ ID NO 18)
GGC AAA GAA ACA GAG TG	F4.2(@379)	(SEQ ID NO 19)
CAG GAC ACA ATG TAG GA	F4.1(@449)	(SEQ ID NO 20)
GTT ATA AAG AAA GAA TTA TAA AG	FJn(@846)	(SEQ ID NO 21)
GAA AAT AAC GCA ATG GAC	F5.1(@875)	(SEQ ID NO 22)

**REVERSE**

GAT GGG ATA CAT CTT TTC C	R6.1(@99)	(SEQ ID NO 23)
CAA GCT ACA TTC AGG TTC CC	F2.2R(@188)	(SEQ ID NO 24)
GGA CTC CAT CGC TCT GCC	R4.1(@510)	(SEQ ID NO 25)
GAC TTA GAA ACT ACT G	R3.4(@694)	(SEQ ID NO 26)
ATA GAC GTG TAA AAC CTG C	R2.1(@735)	(SEQ ID NO 27)
AAC TGT TAT AAA TTT TTA	RSP2(@848)	(SEQ ID NO 28)
CTT TTT CCT TTA TAA TTC TTT C	R2.3o(@875)	(SEQ ID NO 29)

**Figure 34**

### An Additional Splice Product Predicted From The Apo-4 Gene

A second potential theoretical splice product which retains exon 78.3 is shown below.

**H2 p1-124 spliced product =351 bp, 117 amino acids + 10 from vector + 1 N-glycosylation site; predicted weight = 21.9 Kd**

### Figure 35A

#### Peptide Generated

MFVNTTKVEKMYPIMEYSCSDRNLVLIYGILLIYIYIGNLNMKKEQNKCFITPDSRMVFII  
FIQQRGLDSKSLQEINLYFCEGFYTSMLYKKVIRKLHKITQWTRTPQNQSEVEIA (117  
amino acids) (SEQ ID NO 30)

### Figure 35B

Start	Exon No.	Exon Position	Exon Length	Intron No.	Intron Position	Intron Length
@26 bp	78.1	@16-41	26 bp	78.3	@42-74	33 bp
	78.3	@75-181	106 bp	79.1	@182-530	349 bp
	79.1	@531-655	125 bp	79.4	@656-721	66 bp
	79.4	@722-770	49 bp	79.55	@771-876	105 bp
	79.55	@877-894	18 bp	79.75	@895-933	39 bp
	79.85	@934- 967	33 bp			

Hydrophobicity Scale KD; Candidate membrane-spanning segments:

Certain 1 22- 42 1.8833

### Figure 35C

0996264-092801

**Predicted TM structure**

> : Too long to be significative

< : Too short to be significative

U : Loop length

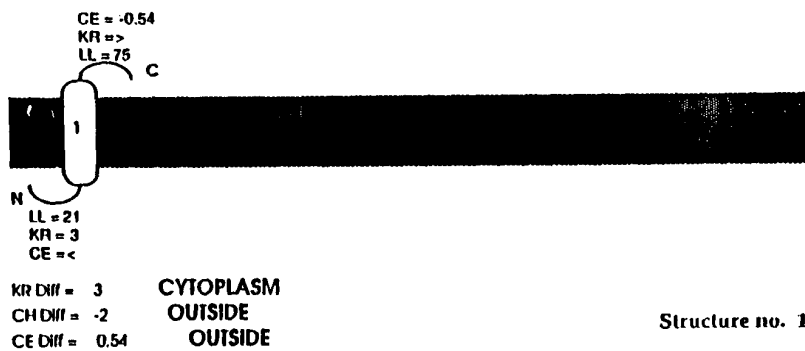
KR : Number of Lys and Arg

KR Diff : Positive charge difference

CE : Net charge energy

CE Diff : Net charge difference

CH Diff : Charge difference over N-term segments



**Figure 35D**



### Nucleic Acid Subsequence Sites Identified In Apo-4

<u>Motif</u>	<u>Position</u>	<u>Significance</u>
CpG	-7, (+28, +106)	DNA methylation site
CAAT	-132, (+127, +131)	Binding of CAAT factors
TATAAT (5/6)	-120, -114, (+10)	TFIID Binding site
TATA	-154	Binds RNA polymerase II and TFIID
CCATTCA	-162, -131	Cap Site I
TATCAGT	+12, (+25)	Cap Site II
TGGCTGCAAGCCCAA (10/14)	-57, (+41)	Binds CTF/NF-I protein
GTGATGG	-140, -4, +11, +32	Eucaryotic Transcription Initiation Site

**Figure 36**

Top Pred predicts 4-5 transmembrane domains for a full-length apo-4F product in which all the stop codons are suppressed.

### Protein sequence and position of predicted TM domains

Begin TM<sub>1</sub>(R)

P1		P2
<p>MFVNTSREKV INQSLIAKVE <u>KMYPIMEYSCSD RNLVLIYGIL LIYIYIGNLN VARHFSMK60</u></p>		
<p>TPVARSNIKL ILTNNVKWLH <u>KKGFASSWKL VKNQTLLCTP SMQLLCCLHP EMGNDFPNGK 120</u></p>		
P3		
<p>ETERCYLSAP FVKSVFLSLC <u>FPGHNVGSLF HMADDLGRAM ESLVSVMTDE EGAEKMFYNS180</u></p>		
<p>RFPHGFYNIH TTKRIRQKEF <u>TRNKSIFLRR VVLYCRFOK FLSLLLFCQ WQVLHVYAIV 240</u></p>		
<p>QKSYKKTCK ILIAKKLAIS <u>LYGTHFGLFK NLKQLKRKNY KGKRKKRNGQ VVKLRTQVCT 300</u></p>		
<p>IIRNTPKPKR GRNSMRSRVR CKLI (324 amino acids) (SEQ ID NO 31)</p>		

Hydrophobicity Scale KD

**Figure 37A**

### Apo-4F : Candidate membrane-spanning segments:

Certain	1	33- 53	1.9073
Putative	2	93- 113	0.8052
Certain	3	124- 144	1.2552
Putative	4	209- 229	1.1833
Putative	5	246- 266	0.9240

I. Transmembrane segments included in structure 8: 1 2 3 4 5; Loop lengths: 32 39 10 64 16 58

**Figure 37B**

K+R difference: -19; -> Orientation: **N-out**; Charge-difference over N-terminal Membr. segs.

(±15 residues): -3; -> Orientation: **N-out**

CYT-EXT profile (neg. values indicate cytoplasmic preference): < < < < -0.13 <

CYT-EXT difference: 0.13

-> Orientation: **N-out**

II. Transmembrane segments included in structure 7: **1 3 4 5**; Loop lengths: 32 70 64 16 58

K+R profile: 5 > 22 > 5; K+R difference: 22 -> Orientation: **N-in**

Charge-difference over N-terminal Membr. segs. (±15 residues): -3; -> Orientation: **N-out**

CYT-EXT profile (neg. values indicate cytoplasmic preference): < -0.13 < -0.26 <

CYT-EXT difference: 0.13; -> Orientation: **N-out**

**Figure 37B (cont'd)**

0996264.092801  
"092801"092801

TopPred predicts a cytoplasmic N-terminus for four TM domains

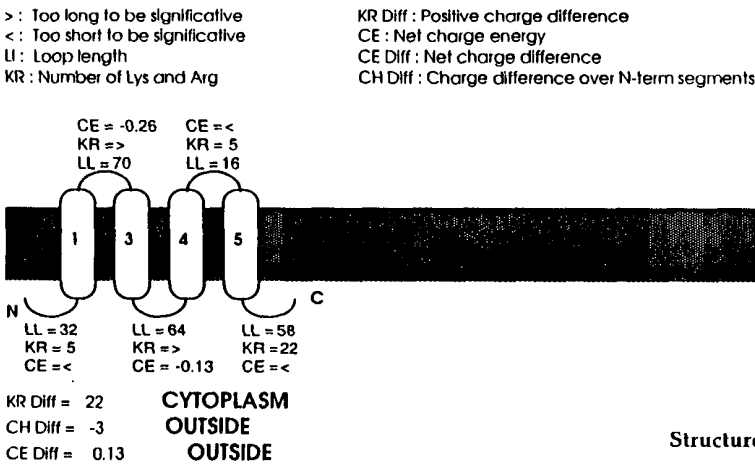
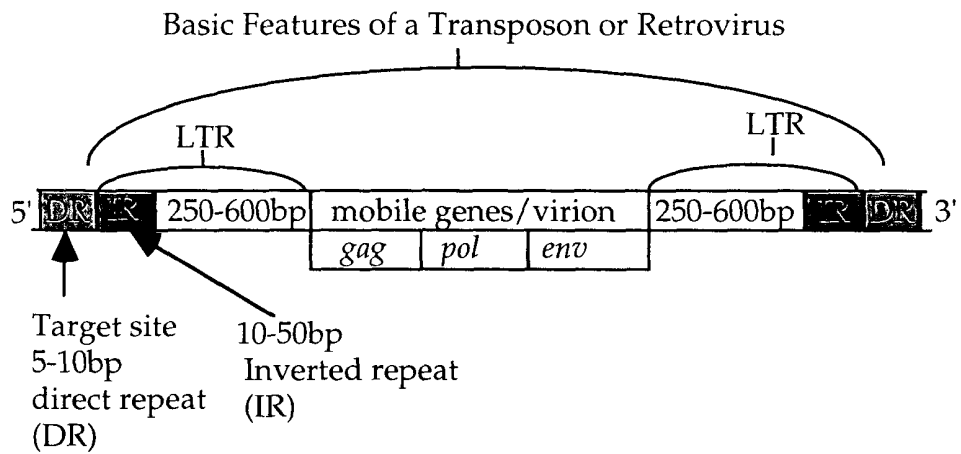
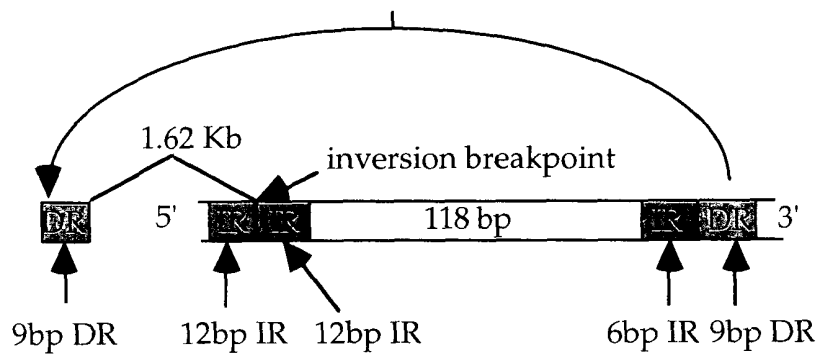


Figure 37C

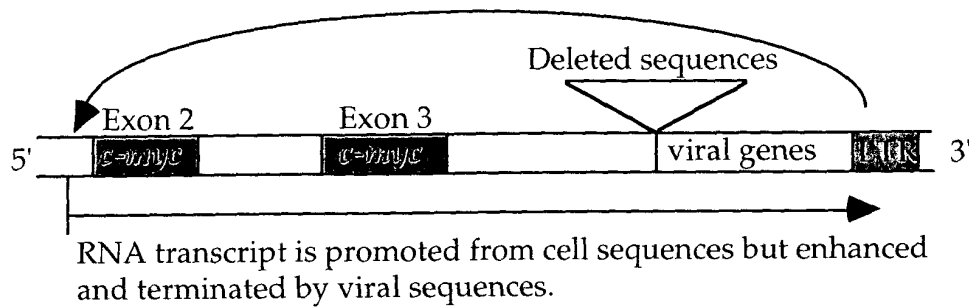


**Figure 38A**

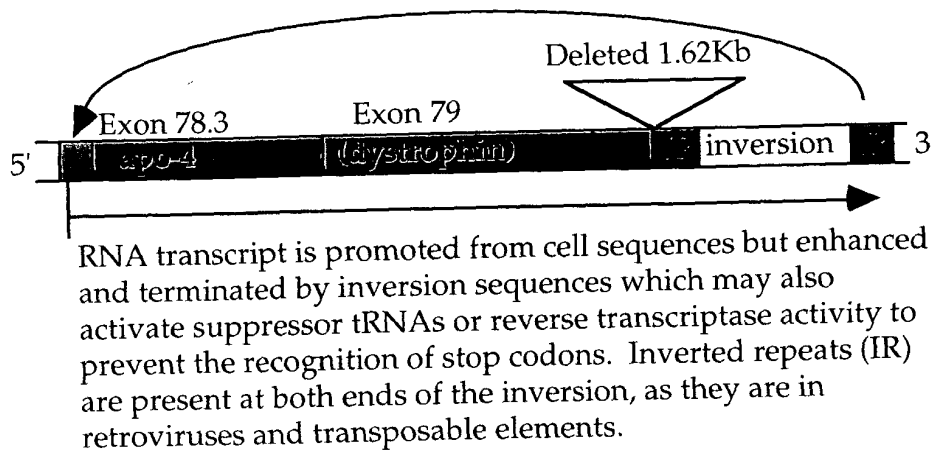
Structure of the apo-4 inversion element before rearrangement



**Figure 38B**



**Figure 39A**



**Figure 39B**